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

(11) PI 0717758-5 B1

(22) Data do Depósito: 10/10/2007

(45) Data de Concessão: 17/11/2020

(54) Título: MÉTODO PARA GERAR UMA PROTEÍNA RUBISCO COM UMA PROPRIEDADE FUNCIONAL MELHORADA

(51) Int.Cl.: C07K 1/00; C07K 14/415; C12N 15/64; C07H 21/04; C12N 15/29.

(30) Prioridade Unionista: 10/10/2006 AU 2006905622.

(73) Titular(es): THE AUSTRALIAN NATIONAL UNIVERSITY.

(72) Inventor(es): JILL E. GREADY; BABU KANNAPPAN.

(86) Pedido PCT: PCT AU2007001542 de 10/10/2007

(87) Publicação PCT: WO 2008/043147 de 17/04/2008

(85) Data do Início da Fase Nacional: 13/04/2009

(57) Resumo: MÉTODO PARA GERAR UMA PROTEÍNA, PROTEÍNA, POLIPEPTÍDEO, POLINUCLEOTÍDEO, CÉLULA HOSPEDEIRA, ORGANISMO FOTOSSINTÉTICO, BEM COMO MÉTODO PARA AUMENTAR SUA EFICIÊNCIA FOTOSSINTÉTICA E MÉTODO PARA AUMENTAR O RENDIMENTO DA COLHEITA OU A RESISTÊNCIA À SECA EM UMA PLANTA. Um método de gerar uma proteína com uma propriedade funcional melhorada, o método compreendendo: (a) identificar pelo menos um Resíduo de aminoácido-alvo em uma primeira proteína, e que o dito Resíduo de aminoácido-alvo é associado à dita propriedade funcional; (b) comparar pelo menos uma segunda proteína homóloga da mesma ramificação filogenética ou uma diferente que a primeira proteína com a primeira proteína e identificar pelo menos um Resíduo de aminoácido Variante entre a primeira proteína e a segunda proteína; (c) selecionar pelo menos um Resíduo de aminoácido Candidato do Resíduo de aminoácido Variante identificado em (b) na base do dito Resíduo de aminoácido Candidato afetando o dito Resíduo de aminoácido-alvo com respeito à dita propriedade funcional; (d) formar pelo menos uma proteína Mutante Candidata in silico ou produzir pelo menos uma proteína Mutante Candidata in vitro em que o dito pelo menos um Resíduo de aminoácido Candidato da segunda proteína substitui um resíduo correspondente na primeira proteína; e (e) tirar a (...).

Relatório Descritivo da Patente de Invenção para **"MÉTODO PARA GERAR UMA PROTEÍNA RUBISCO COM UMA PROPRIEDADE FUNCIONAL MELHORADA".**

REFERÊNCIA CRUZADA AOS PEDIDOS RELACIONADOS

[0001] Este pedido de patente reivindica o benefício do Pedido de Patente Provisório australiano N°. 2006905622 depositado em 10 de outubro de 2006, que é aqui incorporado por referência em sua totalidade.

CAMPO TÉCNICO

[0002] A presente invenção refere-se a um processo para a geração de proteínas com propriedades melhoradas, em particular proteínas Rubisco, e uso das mesmas.

ANTECEDENTES

[0003] Ribulose-1,5-bisfosfato (RuBP) carboxilase/oxigenase é mais comumente conhecida pela abreviação Rubisco. Rubisco é uma enzima que participa na fixação de carbono no Ciclo de Calvin por meio do qual dióxido de carbono atmosférico é fixado e tornado disponível para sistemas biológicos na forma de moléculas ricas em energia.

[0004] Em plantas, algas, cianobactérias e proteobactérias fototrópicas e quimioautotróficas, Rubisco compreende cadeias de subunidade grandes (LSU) e cadeias de subunidade pequenas (SSU). Os sítios de ligação do substrato ficam localizados nas cadeias grandes. As cadeias grandes formam dímeros em que os aminoácidos de cada cadeia grande contribuem para os sítios de ligação. Um total de quatro dímeros de cadeias grandes e oito cadeias pequenas se ajuntam em um complexo maior de cerca de 540.000 Da.

[0005] Rubisco catalisa a primeira etapa na assimilação fotossintética de CO_2 (redução de carbono), como também a fixação competitiva de O_2 que produz o produto residual que é reciclado na oxidação fotor-

respiratória de carbono. A importância de Rubisco é evidenciada pelo fato que é o catalisador principal na reação química fundamental pela qual carbono inorgânico entra nos sistemas biológicos. Além disso, Rubisco é uma proteína muito abundante. Parry et al. (2003) sugere que Rubisco responda por 30-50% de proteína solúvel total em cloroplastos.

[0006] Porém, a abundância relativa de Rubisco pode ser atribuível ao fato que ela é uma enzima de ação muito lenta que apenas fixa algumas moléculas de CO_2 por segundo, em contraste com as milhares de reações químicas por segundo que caracterizam muitas enzimas. A enzima é ineficiente como um catalisador para a carboxilação de RuBP e está sujeita à inibição competitiva por O_2, para inativação por perda de carbamilação, e para inibição sem perspectiva por ligação de RuBP antes da ativação da enzima através de carbamilação por CO_2. Este comportamento não-ótimo torna a taxa de Rubisco limitativa para fotossíntese. Por conseguinte, sob a maioria das condições e quando a luz não estiver, do contrário, limitando a fotossíntese, Rubisco é a enzima limitativa de taxa primária do Ciclo de Calvin.

[0007] Uma vez que Rubisco é frequentemente limitativa de taxa para fotossíntese em plantas, formas melhoradas de Rubisco teriam impacto considerável em aumentar a produtividade agrícola. Várias tentativas foram feitas para aumentar a eficiência das reações mediadas por Rubisco. Abordagens anteriores incluem introduzir construtos que expressam Rubisco de um organismo em outro organismo, aumentando o nível de expressão das subunidades de Rubisco, expressando subunidades pequenas de Rubisco do DNA de cloroplasto, e alterando os genes de Rubisco através de mutagênese para tentar aumentar a especificidade para dióxido de carbono (em oxigênio) ou do contrário aumentar a taxa de fixação de carbono.

[0008] Foram feitas tentativas para introduzir Rubiscos estranhas,

por exemplo, as de algas vermelhas tais como *Galdieria partita* tendo uma especificidade de CO_2/O_2 alta, em plantas verdes floríferas. Esperaría-se melhorar a eficiência fotossintética das plantas de plantação, mas estas tentativas falharam devido a problemas com produção, montagem e regulação da Rubisco estranha na planta hospedeira (Spreitzer e Salvucci, 2002; Parry et al, 2003). Por outro lado, a subunidade grande de Rubisco de tabaco foi de forma bem sucedida substituída com a subunidade grande homóloga da bactéria fotossintética roxa mais simples *Rodospirillum rubrum* que não requer uma subunidade pequena para dobrar-se e organizar-se em uma enzima ativa (Andrews e Whitney, 2003). Embora demonstrando que substituição de Rubisco fosse realizável, a planta transgênica exibiu as propriedades de especificidade e catalíticas muito inferiores de Rubisco de *R. rubrum*. Numerosas tentativas para definir os papéis dos resíduos de sítios ativos em etapas específicas da reação ou modificar e melhorar as propriedades catalíticas de Rubisco foram feitas usando mutagênese locodirigida com observações de estruturas cristalográficas de raio X dos complexos de Rubisco de várias espécies. Porém, estes estudos falharam em fornecer uma definição detalhada e autoconsistente dos vários papéis dos resíduos do sítio ativo no curso de tempo completo da reação. Estas técnicas foram uniformemente mal-sucedidas em engenheirar uma Rubisco "melhor". Embora um mecanismo pelo qual a Rubisco opera possa ser deduzido por estes estudos, pode não ser único devido a possíveis interpretações diferentes dos dados experimentais incompletos, e, desse modo, pode não ser o mecanismo que exista na realidade. Por exemplo, os mecanismos propostos no mecanismo consensos de Cleland para Rubisco assume que uma molécula de água foi deslocada do magnésio no sítio ativo antes da formação do complexo reativo para carboxilação, e que por conseguinte todas as etapas subsequentes na reação também prosseguem na suposição

que este deslocamento aconteça. Porém, não há nenhuma evidência experimental que água é de fato deslocada.

[0009] Isto contrasta com os programas de reengenharia para muitas outras enzimas onde mutações simples, ou em alguns casos mutações múltiplas, que foram deduzidas diretamente dos dados estruturais e mecanísticos obtidos experimentalmente provaram com sucesso em modificar a especificidade do substrato ou eficiência catalítica em direções desejadas previsíveis. Uma dificuldade principal em implementar um método de reengenharia racional para Rubisco é que nenhum dos estudos experimentais relatados dispôs de prova direta da estrutura para todos os intermediários envolvidos, nem os papéis precisos de todos os resíduos de sítio ativo participantes, devido ao estado incompleto acima mencionado dos dados experimentais. Abordagens experimentais não podem inerentemente definir papéis precisos para prótons e moléculas de água envolvidos nos processos catalíticos, visto que eles são "invisíveis" às sondas experimentais. Esta dificuldade é composta pela complexidade tanto do sítio ativo da Rubisco e pelo fato que uma sequência de reações está envolvida. Propõe-se haver vários "elementos" de sítio ativo compreendendo combinações diferentes de resíduos de sítio ativo que tomam parte nas etapas de reações diferentes, com os grupos residuais frequentemente sendo "reusados".

[00010] Parece que as Rubiscos atuais representam apenas "soluções evolutivas parciais" para otimizar a eficiência enzímica, isto é, que processos evolutivos foram incapazes de eficazmente amostrar o espaço da sequência de LSU, e que estas soluções atuais representam soluções longe de ótimas. Desse modo, há uma oportunidade para criar soluções mais ótimas por uma rota diferente, ou combinação de rotas, que evolução biológica tem sido capaz de fornecer até agora.

[00011] A criação ou identificação e introdução de formas mais efi-

cientes de Rubiscos em organismos fotossintéticos mediante transformação, procriação seletiva ou outras manipulações podem permitir crescimento mais eficiente destes organismos, incluindo plantas verdes e em particular plantas floríferas, visto que elas fariam uso mais eficiente de água e nitrogênio, e poderiam crescer mais eficientemente em temperaturas mais altas. Isto por sua vez oferece prospectos para plantações com rendimentos melhores, a revegetação de terra degradada ou propensa à seca, opções melhoradas para sequestro de carbono e melhorias na produção de biocombustível ou energia de biomassa e assim por diante.

[00012] Em resumo, há uma necessidade para gerar proteínas, tais como uma Rubisco, tendo propriedades funcionais melhoradas em que, por exemplo, tais proteínas melhorariam a eficiência e são especificamente adaptadas às condições ambientais particulares.

SUMÁRIO

[00013] Os presentes inventores levantaram a hipótese que as Rubiscos existentes têm provado apenas uma fração do espaço mutacional teoricamente disponível para melhorar sua eficiência e aqueles espaços diferentes podem ter sido provados por grupos diferentes de Rubiscos. Consequentemente, eles propõem o acoplamento destas "soluções" parcial evolutivas enxertando características de mais de um grupo filogenético, ou de mais que uma espécie ambientalmente adaptada, sobre uma Rubisco do hospedeiro para ganhar os benefícios de amostragem mutacional evolutiva mais ampla que foi naturalmente possível. Eles propõem que este processo identificará Rubiscos com eficiência melhorada ou outras propriedades funcionais.

[00014] De acordo com um primeiro aspecto da invenção, é fornecido um método de gerar uma proteína com uma propriedade funcional melhorada, o método compreendendo:

(a) identificar pelo menos um Resíduo de aminoácido-alvo em uma

primeira proteína, em que o dito Resíduo de aminoácido-alvo é associado à dita propriedade funcional;

(b) comparar pelo menos uma segunda proteína homóloga da mesma ou uma ramificação filogenética diferente que a primeira proteína com a primeira proteína e identificar pelo menos um Resíduo de aminoácido Variante entre a primeira proteína e a segunda proteína;

(c) selecionar pelo menos um Resíduo de aminoácido Candidato do Resíduo de aminoácido Variante identificado em (b) na base do dito Resíduo de aminoácido Candidato afetando o dito Resíduo de aminoácido-alvo com respeito à dita propriedade funcional;

(d) formar pelo menos uma proteína Mutante Candidata *in silico* ou produzir pelo menos uma proteína Mutante Candidata *in vitro* em que o dito pelo menos um Resíduo de aminoácido Candidato da segunda proteína substitui um resíduo correspondente na primeira proteína; e

(e) triar a dita pelo menos uma proteína Mutante Candidata produzida em (d) para identificar uma proteína tendo a dita propriedade funcional melhorada. Em uma modalidade, a etapa (a) e a etapa (b) podem ser executadas simultaneamente.

[00015] Em outra modalidade, a etapa (b) pode ser executada antes da etapa (a). Em uma modalidade, a identificação de pelo menos um Resíduo de aminoácido-alvo em uma primeira proteína da etapa (a) pode reduzir a quantidade de espaço da sequência a ser examinada para a identificação de Resíduos de aminoácido Candidatos dos Resíduos de aminoácido Variantes identificados na etapa (b).

[00016] Em uma modalidade, a etapa (d) compreende formar ou produzir pelo menos uma proteína Mutante Candidata usando o pelo menos um Resíduo de aminoácido Candidato da segunda proteína para substituir um resíduo correspondente em uma proteína homóloga e/ou em proteínas homólogas diferentes ou além da primeira proteína.

[00017] Em outra modalidade, a etapa (d) compreende formar ou

produzir pelo menos uma proteína Mutante Candidata em que pelo menos dois Resíduos de aminoácido Candidatos da segunda proteína substituem resíduos correspondentes na primeira proteína e/ou em uma proteína homóloga diferente da primeira proteína.

[00018] Em uma modalidade, o pelo menos um Resíduo de amino-ácido-alvo é contido em uma primeira proteína ou em um conjunto de proteínas contendo a primeira proteína.

[00019] Em certas modalidades, o pelo menos um Resíduo de aminoácido-alvo é pelo menos 2, pelo menos 3, pelo menos 4, pelo menos 5, pelo menos 10, pelo menos 12, pelo menos 15, pelo menos 20, pelo menos 30, ou pelo menos 50 Resíduos de aminoácido-alvos.

[00020] Em uma modalidade, a proteína é uma enzima e a propriedade funcional melhorada é selecionada de qualquer um ou mais dentre eficiência cinética melhorada da enzima, uma especificidade alterada da enzima para um ou mais substratos, uma especificidade alterada para um ou mais produtos da enzima e uma faixa de temperatura eficaz alterada para catálise da enzima. Onde a proteína for Rubisco, a propriedade funcional melhorada pode ser eficiência catalítica melhorada ou especificidade melhorada para dióxido de carbono.

[00021] Em uma modalidade, os Resíduos de aminoácido-alvos são selecionados daqueles resíduos que interagem diretamente com um substrato ou um intermediário da reação, que inclui por exemplo, em Rubisco aqueles resíduos que coordenam diretamente para o centro de reação (resíduos do "primeiro invólucro"), ou coordenando diretamente com um ou mais dos resíduos supracitados, (isto é, resíduos do "segundo invólucro"). Onde a proteína for Rubisco, os resíduos do primeiro invólucro de Rubisco podem ser selecionados de qualquer um ou mais de Glu60, Asn123, Lys175, LYS177, KCX201, Asp203, Glu204, His294 e Lys334.

[00022] Em uma modalidade, os Resíduos de aminoácido-alvos da

proteína de Rubisco estão no domínio N-terminal da subunidade Grande da Rubisco. Os Resíduos de aminoácido-alvos do domínio N-terminal podem estar envolvidos na etapa de adição de gás da catálise da carboxilase mediada por uma enzima de Rubisco. Em uma modalidade particular, os Resíduos de aminoácido-alvos do domínio N-terminal envolvidos na etapa de adição de gás são selecionados do grupo que consiste em ASN123, GLU60 e Tyr20.

[00023] Em uma modalidade, a proteína é uma enzima, e a segunda proteína é selecionada em base das características cinéticas e funcionais adaptadas aos ambientes de crescimento particulares, tais como adaptação ao quente ou ao frio, ou resistência à seca. Estas características podem ser identificadas, por exemplo, dos dados de diversidade ambiental para a segunda proteína.

[00024] Em ainda outra modalidade, o pelo menos um Resíduo de aminoácido Candidato é identificado como um resíduo que é capaz de afetar o dito pelo menos um Resíduo de aminoácido-alvo identificado em (a) e o pelo menos um Resíduo Candidato de amino modula a propriedade funcional da proteína formada. O Resíduo de aminoácido Candidato pode afetar o pelo menos um Resíduo de aminoácido-alvo com respeito à propriedade funcional devido à proximidade do Resíduo de aminoácido Candidato a o pelo menos um Resíduo de aminoácido-alvo. Por exemplo, o efeito pode ser devido a efeitos estéricos, eletrostáticos ou hidrofóbicos.

[00025] Em certas modalidades alternativas, a etapa (c) compreende selecionar pelo menos um Resíduo de aminoácido Candidato Divergente, em vez de pelo menos um Resíduo de aminoácido Candidato, dos Resíduos de aminoácido Variantes identificados em (b). Em modalidades particulares, a etapa (c) compreende selecionar pelo menos 2 Resíduos de aminoácido Candidatos Divergentes.

[00026] Em certas modalidades alternativas, a etapa (c) compreen-

de selecionar pelo menos um Resíduo de aminoácido Candidato Alternativo, em vez de pelo menos um Resíduo de aminoácido Candidato, dos Resíduos de aminoácido Variantes identificados em (b). Em modalidades particulares, a etapa (c) compreende selecionar pelo menos 2 Resíduos de aminoácido Candidatos Alternativos.

[00027] Em outras modalidades particulares, a etapa (c) compreende selecionar pelo menos um Resíduo de aminoácido Candidato Alternativo e pelo menos um Resíduo de aminoácido Candidato Divergente em vez de pelo menos dois Resíduos de aminoácido Candidatos, dos Resíduos de aminoácido Variantes identificados em (b).

[00028] Em uma modalidade, a etapa (a) de identificar pelo menos um Resíduo de aminoácido-alvo compreende a etapa de cálculos de QM de fragmento de sítio ativo e/ou Cálculos de QM/QM híbridas e/ou de QM/MM com base em dados empíricos da estrutura da proteína. Em uma modalidade, os dados empíricos são estrutura de cristal de raio X ou estrutura de RMN de solução. Em uma modalidade, os dados empíricos podem também compreender qualquer um ou mais de dados mutacionais, dados cinéticos, discriminação de isótopos, dados calorimétricos, e espectroscópicos (Fersht, 1998; Frey e Hegeman, 2007). Em uma modalidade, os cálculos de QM/MM são complementados através de simulações de dinâmica molecular (MD) com um potencial de QM/MM (Gready et al, 2006).

[00029] Em uma modalidade, a etapa de selecionar pelo menos um Resíduo de aminoácido Candidato do pelo menos um Resíduo de aminoácido Variante compreende avaliar a proximidade do pelo menos um Resíduo de aminoácido-alvo identificado em (a) aos Resíduos de aminoácido Variantes identificados em (b) e/ou posição relativa às unidades estruturais secundárias.

[00030] A etapa (c) de selecionar pelo menos um Resíduo de aminoácido Candidato pode compreender identificar as alterações entre a

dita proteína e a dita segunda proteína com respeito às interações eletrostáticas e/ou hidrofóbicas do um ou mais Resíduos de aminoácido Variantes identificados em (b) com o pelo menos um Resíduo de aminoácido-alvo identificado em (a) e/ou uma unidade estrutural secundária contendo o pelo menos um Resíduo-alvo identificado na etapa (a). O dito procedimento de seleção também compreende identificação de mutações compensatórias necessárias para remover os efeitos de obstáculos estéricos criados introduzindo um Resíduo de aminoácido Candidato na primeira proteína.

[00031] Em ainda outra modalidade, a etapa de triar o pelo menos um Proteína Mutante Candidata compreende qualquer um ou mais de análise *in silico*, avaliação bioquímica e avaliação fisiológica. A avaliação bioquímica pode incluir avaliação de enovelamento e/ou montagem corretos da proteína, avaliação da estrutura da proteína, avaliação da atividade catalítica e/ou outra função de ligação da proteína ou avaliação da estabilidade da proteína *in vitro*. A avaliação fisiológica pode incluir avaliação de expressão correta, enovelamento e/ou montagem corretos da proteína, avaliação da estrutura da proteína, avaliação da atividade catalítica da proteína ou avaliação da estabilidade da proteína *in vivo*.

[00032] Em ainda outra modalidade do primeiro aspecto, o método também inclui a outra etapa executada após a etapa (c) e antes da etapa (d) de agrupar os Resíduos de aminoácido Candidatos prognosticados como estando usando um efeito cumulativo no dito pelo menos um Resíduo de aminoácido-alvo. Por exemplo, o dito efeito cumulativo pode ser realizado por meio de um ou mais de efeitos eletrostáticos e/ou hidrofóbicos e/ou estéricos, por exemplo, através de efeitos coordenados ou compensatórios no posicionamento das unidades estruturais secundárias e alças, que modificam tais efeitos.

[00033] Em ainda outra modalidade, o método também inclui a ou-

tra etapa executada imediatamente após a etapa (d) de classificar as ditas proteínas Mutantes Candidatas tendo a dita propriedade funcional melhorada. O processo de classificar as ditas proteínas Mutantes Candidatas pode compreender avaliar sua probabilidade de ter a dita propriedade funcional melhorada.

[00034] Este procedimento compreende avaliar as contribuições potenciais relativas dos Resíduos de aminoácido Candidatos e/ou Resíduos de aminoácido Candidatos Alternativos e/ou Resíduos de aminoácido Candidatos Divergentes nas ditas proteínas Mutantes Candidatas para a dita propriedade funcional melhorada. Em uma modalidade, o primeiro aspecto compreende a etapa adicional de usar a informação derivada de pelo menos um ciclo de triar proteínas Mutantes Candidatas na etapa (e) para a identificação de Sub-regiões dentro da estrutura de proteína que preferencialmente influênciam as propriedades dos Resíduos de aminoácido-alvos ligados à dita Região.

[00035] Em outra modalidade, as Sub-regiões podem prover uma base para identificar Resíduos Candidatos, Resíduos Candidatos Alternativos, Resíduos Covariantes e/ou Resíduos Candidatos Divergentes adicionais prognosticados interagir com os Resíduos Candidatos, Resíduos Candidatos Alternativos e/ou Resíduos Candidatos Divergentes identificados através das etapas (a) a (e). Os ciclos de triagem podem ser usados para gerar um conjunto de sítios mutacionais preferidos, e um conjunto preferido de combinações dos mesmos.

[00036] Em ainda outra modalidade do primeiro aspecto, o método compreende a outra etapa de executar evolução direcionada na dita proteína tendo a dita propriedade funcional melhorada e produtos de triagem da mesma.

[00037] Conforme um segundo aspecto da invenção, é fornecido um método de gerar uma proteína de Rubisco com uma propriedade funcional melhorada, o método compreendendo:

(a) identificar pelo menos um Resíduo de aminoácido-alvo em uma primeira proteína de Rubisco, em que o dito Resíduo de aminoácido-alvo é associado à dita propriedade funcional;

(b) comparar pelo menos uma segunda proteína de Rubisco da mesma ramificação filogenética ou uma diferente que a primeira proteína de Rubisco com a primeira proteína de Rubisco e identificar pelo menos um Resíduo de aminoácido Variante entre a primeira proteína de Rubisco e a segunda proteína de Rubisco;

(c) selecionar pelo menos um Resíduo de aminoácido Candidato dos Resíduos de aminoácido Variantes identificados em (b) na base do dito Resíduo de aminoácido Candidato afetando o dito Resíduo de aminoácido-alvo com respeito à dita propriedade funcional;

(d) formar pelo menos uma proteína de Rubisco Mutante *in silico* ou produzir pelo menos uma proteína de Rubisco Mutante *in vitro* em que o dito pelo menos um Resíduo de aminoácido Candidato da segunda proteína de Rubisco substitui um resíduo correspondente na primeira proteína de Rubisco; e

(e) triar a dita pelo menos uma proteína Mutante Candidata de Rubisco produzida em (d) para identificar uma proteína de Rubisco tendo a dita propriedade funcional melhorada.

[00038] Em uma modalidade, a etapa (a) e a etapa (b) podem ser executadas simultaneamente.

[00039] Em outra modalidade, a etapa (b) pode ser executada antes da etapa (a).

[00040] Em outra modalidade, a identificação de pelo menos um Resíduo de aminoácido-alvo em uma primeira proteína da etapa (a) reduz a quantidade de espaço da sequência a ser examinada para a identificação de Resíduos de aminoácido Candidatos dos Resíduos de aminoácido Variantes identificados na etapa (b).

[00041] Em outra modalidade do segundo aspecto, os resíduos da

segunda proteína de Rubisco usados para comparação com a primeira proteína de Rubisco compreendem todos os resíduos diretamente coordenados para o sítio ativo de uma proteína de Rubisco, todos os resíduos que interagem com o centro reativo do substrato ou espécies de reação intermediárias de uma proteína de Rubisco, e outros resíduos dentro de uma distância aproximada do sítio ativo de uma proteína de Rubisco ou um subconjunto de tais resíduos. por exemplo, a distância aproximada pode ser entre 3 e 28 Å de qualquer átomo de substrato ou espécies de reação intermediárias. Tipicamente, a distância aproximada é entre 6 e 20 Å de qualquer átomo de substrato ou espécies de reação intermediárias. Mais tipicamente, a distância aproximada é entre 9 e 15 Å de qualquer átomo de substrato ou espécies de reação intermediárias.

[00042] Em uma modalidade do segundo aspecto, a primeira proteína de Rubisco é tirada de espécies de plantas verdes e cianobactérias e a segunda proteína de Rubisco é tirada de espécies de algas vermelhas. Em uma modalidade particular, a primeira proteína de Rubisco é tirada de espécies de plantas floríferas e cianobactérias e a segunda proteína de Rubisco é tirada de espécies de algas vermelhas.

[00043] Em uma modalidade do segundo aspecto, a etapa (c) compreende selecionar pelo menos um Resíduo de aminoácido Candidato Divergente, em vez de pelo menos um Resíduo aminoácido de Candidato, dos Resíduos de aminoácido Variantes identificados em (b). Em modalidades particulares, a etapa (c) compreende selecionar pelo menos 2 Resíduos de aminoácido Candidatos Divergentes

[00044] Em certas modalidades alternativas do segundo aspecto, a etapa (c) compreende selecionar pelo menos um Resíduo de aminoácido Candidato Alternativo, em vez de pelo menos um Resíduo de aminoácido Candidato, dos Resíduos de aminoácido Variantes identificados em (b). Em modalidades particulares, a etapa (c) compreende

selecionar pelo menos 2 Resíduos de aminoácido Candidatos Alternativos.

[00045] Em outras modalidades particulares, a etapa (c) compreende selecionar pelo menos um Resíduo de aminoácido Candidato Alternativo e pelo menos um Resíduo de aminoácido Candidato Divergente em vez de pelo menos dois Resíduos de aminoácido Candidatos, dos Resíduos de aminoácido Variantes identificados em (b). Em uma modalidade do segundo aspecto, é fornecido um método de purificar a proteína de Rubisco, o método que compreende as etapas de:

(a) fundir em um primeiro vetor a sequência de codificação com uma Sequência de ubiquitina (Ub) marcada com H_6 (H_6Ub) para a extremidade de 5' de um gene *rbc*S;

(b) cotransformar o primeiro vetor com um segundo vetor codificando para a subunidade grande e subunidade pequena da proteína de Rubisco em um hospedeiro;

(c) induzir expressão da dita proteína de Rubisco e vetores;

(d) purificar a proteína de Rubisco com base na expressão do marcador de ubiquitina fundido à subunidade pequena de Rubisco;

(e) remover fragmentos de Ub da Rubisco.

[00046] Em uma modalidade, a etapa (d) de purificar a dita proteína é executada usando cromatografia tal como cromatografia de afinidade de metal.

[00047] Em uma modalidade, o primeiro e/ou segundo vetor é um plasmídeo.

[00048] Em uma modalidade, o hospedeiro é *E. coli*.

[00049] Em uma modalidade, a subunidade grande compreende uma ou mais mutações.

[00050] Em uma modalidade, os fragmentos de Ub são removidos usando uma protease Ub-específica.

[00051] De acordo com um terceiro aspecto da invenção, é forneci-

da uma proteína produzida pelo método do primeiro ou segundo aspectos da invenção.

[00052] De acordo com um quarto aspecto da invenção, é fornecida uma proteína de Rubisco produzida pelo método do primeiro ou segundo aspectos da invenção.

[00053] De acordo com um quinto aspecto da invenção, é fornecida uma proteína de Rubisco compreendendo a sequência como exposta em SEQ ID NOS: 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 74, 76, 78, ou um equivalente funcional das mesmas.

[00054] De acordo com um sexto aspecto da invenção, é fornecida uma proteína de Rubisco codificada por um polinucleotídeo que compreende a sequência exposta em SEQ ID NOS: 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 73, 75, 77, ou um equivalente funcional das mesmas.

[00055] De acordo com um sétimo aspecto da invenção, é fornecido um polipeptídeo de subunidade grande de Rubisco que compreende uma substituição de resíduo de aminoácido ou uma combinação de substituições de resíduos de aminoácido selecionadas do grupo que consiste em (Y25W, D51I), (Y25W, D51V), (T23G, K81R), (G54A, C84A, I87V), (G54S, C84A, I87V), (T23G, Y25W, D51I, K81R), (Y25W, D51I, G54A, C84A, I87V), (Y25W, D51I, G54S, C84A, I87V), (Y25W, D51V, G54A, C84A, I87V), (V121I, M297G, V300T), (L36I, I116L, F140L), (L36I, I116L, V121I, F140L, M297G, V300T), (KI8I, T23G), (K21A, L22K, (intervalo)M, T23G, Y25W), (T23G, K18I, T68V, K81R), (T23G, K81R, P104E), (T23G, D19P, K81R), (T23G, K81R, V121I, M297G, V300T), (T23G), (K81R), (V121I, M297G), (M297G), (V1221I).

[00056] Em uma modalidade, a subunidade grande de Rubisco está presente em uma proteína de Rubisco. Em uma modalidade, a Rubisco é fornecida na forma de uma proteína de fusão ou um fragmento

que retêm a atividade biológica da Rubisco.

[00057] Também fornecido é um polinucleotídeo que codifica um polipeptídeo de LSU de Rubisco de acordo com o sétimo aspecto.

[00058] Também fornecido é um vetor que compreende a sequência de polinucleotídeo do quinto aspecto ou a sequência de polinucleotídeo definida no sexto aspecto.

[00059] Em uma modalidade o vetor compreende um promotor de expressão constitutivo ou selecionável. Em uma modalidade, o vetor compreende um marcador selecionável.

[00060] De acordo com um oitavo aspecto da invenção, é fornecida uma célula hospedeira transformada com uma sequência de ácido nucleico ou o vetor de acordo com os aspectos acima.

[00061] Em uma modalidade, a célula hospedeira é uma célula procariótica ou uma célula eucariótica. Em certas modalidades, a célula hospedeira procariótica é uma célula bacteriana tal como *E. coli*.

[00062] De acordo com um nono aspecto da invenção, é fornecido um organismo fotossintético transformado com uma sequência de ácido nucleico ou um vetor de um dos aspectos acima. Em uma modalidade, o organismo fotossintético é uma cianobactéria.

[00063] Em outra modalidade, o organismo fotossintético é um *Synechococcus*, tal como PCC7942 de *Synechococcus* sp. ou PCC6301 de *Synechococcus* sp.

[00064] Em uma modalidade o organismo fotossintético é uma planta florífera. A planta florífera pode ser um tabaco do tipo selvagem ou transgênico (*Nicotiana tabacuni*). O tabaco transgênico pode ser um tabaco em que *rbc*L de tabaco nativo foi substituído por *rbc*M de *Rodospirillum rubrum*.

[00065] Em uma modalidade, a proteína é expressada em organelas fotossintéticas.

[00066] Em ainda outra modalidade, a organela é um plastídio.

[00067] Em ainda outra modalidade, a organela pode ser selecionada do grupo de plastídios de eucariotos fotossintéticos compreendendo cloroplasto (plastídio contendo clorofila), etioplasto (cloroplasto não exposto à luz), cromoplasto (plastídio não contendo clorofila), ou leucoplasto (plastídios não-pigmentados para armazenar amido (amiloplasto), lipídio (elaioplasto) ou proteína (proteinoplasto)).

[00068] De acordo com um décimo aspecto da invenção, é fornecido um método de aumentar a eficiência fotossintética de um organismo, o método compreendendo introduzir uma sequência de ácido nucleico que codifica uma proteína de Rubisco de acordo com qualquer um dos aspectos acima no dito organismo e expressar a proteína de Rubisco.

[00069] De acordo com um décimo primeiro aspecto da invenção, é fornecido um método de aumentar o rendimento da colheita, o método compreendendo introduzir uma sequência de ácido nucleico que codifica uma proteína de Rubisco de um dos aspectos acima na dita planta de plantação.

[00070] De acordo com um décimo segundo aspecto da invenção, é fornecido um método de aumentar a resistência à seca em uma planta em que o dito método compreende introduzir uma sequência de ácido nucleico que codifica uma proteína de Rubisco de acordo com um dos aspectos acima na dita planta.

[00071] De acordo com um décimo terceiro aspecto da invenção, é fornecido um método de aumentar a biomassa em uma ou mais plantas ou outro(s) organismo(s) fotossintético(s), em que o dito método compreende introduzir uma sequência de ácido nucleico que codifica uma proteína de Rubisco de acordo com um dos aspectos acima na(s) dita(s) planta(s) ou organismo(s).

[00072] De acordo com um décimo quarto aspecto da invenção, é fornecido um método de produzir um biocombustível que compreende

material de uma planta ou plantas ou outro(s) organismo(s) fotossintético(s), em que o dito método compreende introduzir uma sequência de ácido nucleico que codifica uma proteína de Rubisco de acordo com um dos aspectos acima na(s) dita(s) planta(s) ou organismo(s).

[00073] Em uma modalidade do décimo ao décimo quarto aspecto, a introdução de proteínas pode compreender a etapa de transformação, reprodução sexual, ou uma combinação destes.

ABREVIAÇÕES

aada	gene de plasmídeo conferindo resistência à estreptomicina e espectinomicina
ACR	Resíduo Candidato Alternativo
2CABP	1,5-bisfosfato de 2-carboxiarabinitol
2C3KABP	1,5-bisfosfato de 2-carbóxi-3-cetoarabinitol
3-PGA	3-fosfo-D-glicerato
CM	Mutante Candidato
CPK	representação molecular de enchimento espacial de Corey-Pauling-Koltun dos átomos esféricos de raios proporcionais ao raio de van der Walls dos átomos
CR	Resíduo Candidato
CvR	Resíduo Covariante
DCR	Resíduo Candidato Divergente
DFT	teoria funcional de densidade
ESP	potencial eletrostático
FM	fragmento de modelo
H_6	6x marcador de afinidade de histidina
K_c	constante de Michaelis (K_m) para CO_2 a 0% oxigênio
K_c^{ar}	constante de Michaelis (K_m) para CO_2 a 21% (ambiente) oxigênio
k_c^{cat}	taxa de giro de carboxilação
K_0 ou K_{10}	constante de Michaelis (K_m) para O_2

LSU	subunidade grande
MD	dinâmica molecular
código PDB	número de identidade do Banco de Dados de Proteínas
ONIOM	Nosso próprio Método de Mecânica orbital mocular + molecular integrada em n-camadas
QM/MM	mecânica quântica
QM/MM	mecânica quântica híbrida/mecânica molecular
QM/QM	mecânica quântica híbrida/mecânica quântica
*rbc*L	polinucleotídeo codificando uma subunidade grande de Rubisco
*rbc*LS	polinucleotídeo codificando uma subunidade grande de Rubisco de *Synechococcus* sp.
*rbc*LS-*rbc*SS (ou *rbc*LS)	polinucleotídeo codificando uma subunidade grande de Rubisco e subunidade pequena de Rubisco de *Synechococcus* sp.
*rbc*M	polinucleotídeo codificando a Rubisco de Forma II de *Rodospirillum rubrum*
*rbc*S	polinucleotídeo codificando uma subunidade pequena de Rubisco
*rbc*SS	polinucleotídeo codificando uma subunidade pequena de Rubisco de *Synechococcus* sp.
RuBP	Ribulose-1,5-bisfosfato
Rubisco	Ribulose-1,5-bisfosfato carboxilase/oxigenase
$S_{c/o}$ definido	especificidade da Rubisco para CO_2 comparada a O_2 como $k_c^{cat}/K_c)/(k^0_{cat}/K_o)$
SSU	subunidade pequena
SsVR	Resíduo Variante Espécie-específico
TIM barrel	uma estrutura do domínio de proteína definida primeiro em triosefosfato isomerase

TR	Resíduo-alvo
TS	estado de transição
Ub	ubiquitina
V_c^{max}	atividade da carboxilase máxima extrapolada
VR	Resíduo Variante

DEFINIÇÕES

[00074] Como usado neste pedido de patente, a forma singular "um", "uma" e "o/a" inclui referências plurais a menos que o contexto dite claramente diferente. Por exemplo, o termo "uma célula vegetal" também inclui uma pluralidade de células vegetais.

[00075] Como aqui usado, o termo "compreendendo" significa "incluindo". Variações da palavra "compreendendo", tais como "compreendem" e "compreende", têm significados correspondentemente variados. Desse modo, por exemplo, um polinucleotídeo "compreendendo" uma sequência que codifica uma proteína pode consistir exclusivamente naquela sequência ou pode incluir uma ou mais sequências adicionais.

[00076] Por "célula hospedeira" é significado uma célula contendo uma construção de ácido nucleico introduzida e suporta a replicação e/ou expressão da construção. Células hospedeiras podem ser células procarióticas tais como *E. coli*, ou células eucarióticas tais como algas, fungos, levedura, inseto, anfíbio, nematódeo, planta ou células mamíferas. As células hospedeiras podem ser células vegetais, tais como células vegetais monocotiledôneas ou células vegetais dicotiledôneas. Um exemplo de uma célula hospedeira é uma célula hospedeira de *E. coli*.

[00077] O termo "planta verde" como aqui usado é intencionado abranger organismos incluindo, mas não necessariamente limitados a, organismos unicelulares ou multicelulares das Divisões *Pteridófitas* (samambaias), *Briófitas* (musgos), *Carófitas* e *Clorófitas* (algas verdes

aquáticas), *Magnoliófitas* (plantas floríferas ou angiospermas), e *Pinófitas* (coníferas).

[00078] Como aqui usado, proteínas "homólogas" são proteínas que compartilham uma origem evolutiva. Proteínas homólogas podem compartilhar a mesma função essencial (proteínas ortólogas) ou podem apresentar funções significativamente diferentes, evolucionariamente divergentes (proteínas parálogas).

[00079] Como aqui usado, "ácido nucleico" significa um polinucleotídeo e inclui polímero uni ou bifilamentar de bases de deoxirribonucleotídeo ou de ribonucleotídeo. Ácidos nucleicos podem também incluir fragmentos e nucleotídeos modificados.

[00080] Como aqui usado "operavelmente ligado" inclui referência a uma ligação funcional de pelo menos duas sequências. Operavelmente ligado inclui ligação entre um promotor e uma segunda sequência, em que a sequência de promotor inicia e medeia a transcrição da sequência de DNA que corresponde à segunda sequência. O *rbc*L-S é um exemplo de um sequência operavelmente ligada.

[00081] Como aqui usado, "fotossíntese" refere-se ao processo em plantas verdes e certos outros organismos pelos quais carboidratos são sintetizados de dióxido de carbono e água usando luz como uma fonte de energia. A maioria das formas de fotossíntese libera oxigênio como um subproduto.

[00082] Como aqui usado, "ramificação filogenética" refere-se a um grupo de organismos conectados por linhagem. No contexto da presente invenção, filogenéticos constituem um meio de classificar grupos de organismos de acordo com grau de parentesco evolutivo. Uma ramificação filogenética pode conter organismos de diferentes divisões taxonômicas, classe, ordem, família, gênero ou espécies.

[00083] "Enxerto filogenético" refere-se ao processo de introduzir pelo menos um resíduo de aminoácido de uma proteína doadora de

um organismo de uma ramificação filogenética em uma proteína recipiente de um organismo de uma ramificação filogenética diferente para o propósito de melhorar as propriedades funcionais da proteína recipiente. Enxerto filogenético pode ser realizado substituindo pelo menos um resíduo de aminoácido em uma posição dada na sequência de uma proteína recipiente, o pelo menos um resíduo de aminoácido estando na mesma posição na segunda proteína doadora selecionada em base da análise filogenética. Como aqui usado, "planta" inclui plantas e partes de planta incluindo, mas não limitadas às células vegetais, tecido de planta tais como folhas, talos, raízes, flores, e sementes.

[00084] Como aqui usado, "o promotor" inclui referência a uma região de DNA que está envolvida no reconhecimento e ligação de uma RNA polimerase e outras proteínas para iniciar a transcrição.

[00085] Como aqui usado, "proteína" refere-se a qualquer polímero de aminoácidos ligado através de ligações peptídicas ou ligações peptídicas modificadas, ou natural ou sinteticamente produzidas. A proteína da invenção pode compreender componentes não-peptídicos, tais como grupos carboidrato. Carboidratos e outros substituintes não-peptíticos podem ser adicionados a uma proteína pela célula na qual a proteína é produzida, e varia com o tipo de célula.

[00086] Proteínas são definidas aqui, em termos de suas estruturas de cadeia principal de aminoácido; substituintes tais como grupos de carboidratos são em geral não especificados, mas podem estar, no entanto, presentes.

[00087] Como definido aqui, a "função" de uma proteína refere-se ao papel normal para a proteína na célula, ou um papel para o qual uma proteína pode ser engenheirada a realizar. Em uma modalidade, a proteína é uma enzima. Onde a proteína for uma enzima, a função pode ser a catálise de pelo menos uma reação química. Em outras modalidades a função pode ser estrutural (por exemplo, servindo como

uma proteína citoesquelética). A função pode envolver o transporte ativo ou passivo de uma substância dentro da célula ou entre o interior e exterior da célula, ou entre compartimentos diferentes dentro da célula, ou entre regiões diferentes do organismo, por exemplo, onde a proteína estiver envolvida em um canal ou um poro da membrana, ou a proteína estiver envolvida em tráfego de materiais para compartimentos celulares específicos ou a proteína atuar como um chaperona ou um transportador. A função pode estar envolvida com interações de ligando/receptor, por exemplo, onde a proteína for um fator de crescimento, uma citocina, um neurotransmissor ou um ligante intracelular ou extracelular, ou a proteína for um receptor para o fator de crescimento, citocina, neurotransmissor ou o ligante intracelular ou extracelular.

[00088] Onde a proteína for uma enzima, a enzima pode estar envolvida em catabolismo ou metabolismo. A enzima pode estar envolvida na síntese de pelo menos um produto. A enzima pode estar envolvida no desarranjo de pelo menos um substrato. A enzima pode estar envolvida na modificação química de pelo menos um substrato, por exemplo, a adição ou deleção de um ou mais grupos de fosfato de uma molécula. As enzimas podem estar envolvidas na degradação de pelo menos um substrato.

[00089] Desse modo, uma "propriedade funcional" de uma proteína é uma propriedade que contribui para a função da proteína. por exemplo, onde uma proteína for uma enzima, a propriedade funcional pode ser a especificidade da enzima para um substrato particular, a eficiência cinética da enzima, sua faixa de temperatura eficaz para catálise, ou sua especificidade para catalisar sua reação normal e minimizar reações colaterais para subprodutos indesejados e/ou potencialmente tóxicos.

[00090] O termo "resíduo" no contexto de um polipeptídeo refere-se

a uma unidade de aminoácido na cadeia de polipeptídeo linear. É o resta de cada aminoácido, i.e -NH-CHR-C-, após água ser removida na formação do polipeptídeo de α-aminoácidos, isto é, NH_2-CHR-COOH.

[00091] Os termos "Resíduo de aminoácido-alvo" ou "Resíduo-alvo" (TR) referem-se a um resíduo de aminoácido que é identificado e/ou prognosticado contribuir diretamente para a função da proteína. Onde a proteína for uma enzima, a contribuição de um Resíduo-alvo para a propriedade funcional pode ser uma influência direta do Resíduo-alvo na(s) reação(ões) catalítica(s) realizada(s) pela enzima no uma ou mais do(s) sítio(s) ativo(s) da enzima, por meio de interação direta ou envolvimento no sítio ativo. Um Resíduo-alvo não ficará distante de um sítio ativo da enzima. Onde a proteína for um receptor, o Resíduo-alvo estará envolvido diretamente no sítio de receptor. Nos abordagens descritos aqui, um Resíduo-alvo pode ser identificado por abordagens computacionais e/ou abordagens de simulação molecular, visto que estes abordagens podem considerar as posições das moléculas da água, prótons, estados de ionização e ligações de hidrogênio associadas aos Resíduos-alvos no sítio ativo da enzima, ou sítio do receptor que não é unambiguamente definível por experimento. Nos abordagens da presente invenção, é antecipado que a mutação dos Resíduos-alvos em geral conduzirá a um rompimento ou redução da função. Consequentemente nos abordagens da presente invenção, os Resíduos-alvos não são alterados diretamente através da substituição com outro aminoácido, mas do contrário as propriedades dos Resíduos-alvos tais como sua posição e carga são "sintonizadas" pela manipulação de um ou mais resíduos que interagem com os Resíduos-alvos.

[00092] Os termos "Resíduo de aminoácido Variante" ou "Resíduo Variante" (VR) referem-se a um resíduo de aminoácido específico de

uma segunda proteína ou um resíduo de aminoácido específico identificado em uma sequência de consenso de uma pluralidade das segundas proteínas que é identificado como sendo diferente do resíduo de aminoácido correspondente encontrado em uma primeira proteína que é homóloga à segunda proteína. Resíduos variantes podem ser identificados, por exemplo, usando um alinhamento das sequências de aminoácido das primeira e segunda proteínas. As sequências das primeira e/ou segunda proteínas podem ser sequências de consenso. As sequências podem ser derivadas de organismos do mesmo ou de uma ramificação filogenética diferente. Para a subunidade grande de Rubisco, é raro haver adições de sequência ou deleções entre as sequências de organismos diferentes, além de nas regiões N ou C-terminais que não são importantes para a função catalítica da Rubisco. Não obstante, em uma modalidade um Resíduo de aminoácido Variante pode ser um resíduo que esteja presente em uma primeira proteína para o propósito da presente invenção, mas ausente em uma segunda proteína, ou que esteja ausente em uma primeira proteína mas presente em uma segunda proteína.

[00093] Os termos "Resíduo de aminoácido Candidato" ou "Resíduo Candidato" (CR) referem-se a um resíduo de aminoácido de uma segunda proteína da qual é selecionado dentre uma pluralidade Resíduos variantes e que é suspeito de ser capaz de influenciar um ou mais Resíduos-alvos estérica e/ou eletrostaticamente, e assim influenciar a função da proteína mediada pelo um ou mais Resíduos-alvos. Um Resíduo Candidato é um resíduo que pode ser seletivamente transplantado em uma primeira proteína de um hospedeiro para tentar modular a atividade funcional da primeira proteína para a atividade funcional desejada da segunda proteína.

[00094] Em algumas modalidades, o Resíduo Candidato pode ser selecionado em base da semelhança e/ou diferença dos resíduos de

aminoácido entre duas ou mais ramificações filogenéticas. No contexto de selecionar Resíduos Candidatos para Rubisco, um Resíduo Candidato é um que é suspeito de ser capaz de influenciar um ou mais Resíduos-alvos estérica e/ou eletrostaticamente, por exemplo, modificando a distribuição de carga no Resíduo-alvo e/ou modificando a posição espacial do Resíduo-alvo e/ou modificando a habilidade do Resíduo-alvo para mover-se.

[00095] Onde a proteína for Rubisco, o Resíduo Candidato pode estar presente na sequência de consenso de algas vermelhas mas diferente na posição do resíduo correspondente nas sequências de consenso de plantas floríferas e cianobactérias, e em uma posição onde o aminoácido for o mesmo em plantas floríferas e cianobactérias. Em modalidades particulares, o Resíduo de aminoácido Candidato pode ser selecionado do resíduo em algas vermelhas.

[00096] Onde a proteína for Rubisco, identificação de um Resíduo Candidato como suspeito de afetar a etapa de adição de gás permite o mesmo ser diferenciado dos outros Resíduos Variantes mostrando alterações não-conservadas entre as ramificações ou sub-ramificação, que pode representar desvio filogenético neutro, ou que pode ter um papel fisiológico específico para ramificação, tal como enovelamento, montagem, incluindo interações com a subunidade pequena, ou estabilidade.

[00097] Os termos "Resíduo de aminoácido Candidato Divergente" ou "Resíduo Candidato Divergente" (DCR) referem-se a um resíduo de aminoácido que é selecionado dentre uma pluralidade de Resíduos Variantes e que é suspeito de ser capaz de influenciar um ou mais Resíduos-alvos estérica e/ou eletrostaticamente, e assim influenciar a função da proteína mediada pelo um ou mais Resíduos-alvos. O Resíduo Candidato Divergente é selecionado em base da diferença de resíduos de aminoácido em uma posição dada na sequência de consen-

so da proteína entre pelo menos três ramificações filogenéticas. Por exemplo, no contexto de selecionar Resíduos Candidatos Divergentes para Rubisco, um Resíduo Candidato Divergente é um que é suspeito de ser capaz de influenciar um ou mais Resíduos de aminoácido-alvos de Rubisco estérica e/ou eletrostaticamente, e que pode estar presente na sequência de consenso de algas vermelhas mas diferente no resíduo correspondente da sequência de consenso de plantas floríferas e a sequência de consenso de cianobactérias, e também diferente no resíduo correspondente da sequência de consenso de plantas floríferas e a sequência de consenso de cianobactérias.

[00098] O Resíduo Candidato Divergente pode ser selecionado da sequência ou sequência de consenso da proteína de qualquer uma das ramificações filogenéticas que foram comparadas.

[00099] Os termos "Resíduo de aminoácido Candidato Alternativo" ou "resíduo Candidato Alternativo" (ACR) referem-se a um aminoácido alternativo na posição de um Resíduo Candidato que é expressado em uma segunda proteína, mas que não é o aminoácido que é expressado na sequência de consenso da segunda proteína. Desse modo um Resíduo Candidato Alternativo fornece um resíduo que é expressado na posição dada em pelo menos uma espécie da segunda proteína mas que não é expressado na maior parte das sequências da mesma ramificação filogenética. Onde a proteína for Rubisco, um ACR pode ser selecionado, por exemplo, da sequência de Rubisco de *Griffithsia monolis* que exibe uma taxa catalítica significativamente mais alta que uma Rubisco de uma espécie algal vermelha típica.

[000100] Os termos "Resíduo de aminoácido de Covariante" ou "Resíduo Covariante" (CvR) referem-se a um resíduo que é identificado na sequência de uma segunda proteína de uma espécie particular, como estando na redondeza de um Resíduo Candidato Alternativo e mostrando variação complementar ao Resíduo Candidato Alternativo. Esta

variação do Resíduo Covariante que não está presente na sequência de consenso para a segunda proteína pode ser suspeita de refletir as alterações complementares nas propriedades estruturais e/ou eletrostáticas do Resíduo Candidato Alternativo e Resíduo Covariante. Identificação de Resíduos Covariantes fornece um meio para identificar as posições dos resíduos auxiliares que podem ser mutadas na primeira proteína para melhor acomodar as alterações feitas de transferir os Resíduos Candidatos Alternativos.

[000101] Os termos "Resíduo de aminoácido Variante Espécie-específico" ou "Resíduo Variante Espécie-específico" (SsVR) referem-se aos resíduos que variam entre espécies estritamente relacionadas de uma proteína. Junto com os dados funcionais associados para cada proteína variante, SsVRs podem ser usados para mapear as relações de sequência-estrutura-função. Estas correlações podem ser usadas para prognosticar quais resíduos variáveis poderiam estar contribuindo mais para a melhoria na propriedade desejada para a proteína. Esta informação de SsVR pode ser usada junto com o método geral para identificar CRs. Predições de uma proteína Mutante Candidata contendo grupos de resíduos podem incluir SsVRs além dos CRs, ACRs, DCRs ou CvRs que mais diretamente afetam os Resíduos-alvos.

[000102] Os termos "proteína Mutante Candidata" ou "Mutante Candidato" (CM) refere-se a uma proteína mutante em que pelo menos um ou grupos de dois ou mais CRs e/ou ACRs e/ou DCRs são combinados, com CvRs ou SsVRs adicionais opcionais, em uma proteína simples.

[000103] O termo "Região" refere-se a uma divisão da estrutura de proteína que circunda os Resíduos-alvos que pode ser feita em base da proximidade de um Resíduo-alvo particular ou Resíduos-alvos para partes particulares do centro reativo ou sítio de ligação. Uma Região compreende um volume espacialmente contíguo da estrutura de prote-

ína contendo subconjuntos de CRs, DCRs, ACRs, CvRs e SsVRs que podem influenciar as interações de um Resíduo-alvo particular ou Resíduos-alvos preferencialmente com partes particulares do centro reativo ou sítio de ligação. Os limites das Regiões não são precisamente definidos. As Regiões podem ter segmentos sobrepostos da estrutura da proteína. O propósito de definir Regiões é facilitar a aplicação do método de enxerto filogenético livremente identificando os subconjuntos de CRs, DCRs, ACRs, CvRs e SsVRs que podem ser agrupados para formar Mutantes Candidatos. Por exemplo, onde a proteína for Rubisco, uma região rotulada Região 1, que compreende a maioria do domínio N-terminal de subunidade grande e vários segmentos pequenos da estrutura do domínio C-terminal da subunidade grande adjacente, pode ser identificada como podendo influenciar Asn123, Glu60 e Tyr20, que especificamente interagem direta ou indiretamente com os Resíduos-alvos, com o grupo carboxilato nascente das espécies reativas durante a etapa de adição gasosa.

[000104] O termo "Sub-região" refere-se a uma divisão de uma Região que pode ser feita em base da proximidade de um Resíduo-alvo particular ou Resíduos-alvos a uma parte do componente da Região. Uma Sub-região compreende um volume espacialmente contíguo da estrutura da proteína contendo um subconjunto dos CRs, DCRs, ACRs, CvRs e SsVRs da Região que são prognosticados influenciar as propriedades do Resíduo-alvo particular ou Resíduos-alvos preferencialmente ligados à Região. Os limites das Sub-regiões não são precisamente definidos. As Sub-regiões podem ter segmentos sobrepostos da estrutura da proteína. Definindo Sub-regiões pode facilitar a aplicação do método de enxerto filogenético livremente identificando os subconjuntos dos CRs, DCRs, ACRs, CvRs e SsVRs da Região que podem ser preferencialmente agrupados para formar Mutantes Candidatos. Por exemplo, onde a proteína for Rubisco, regiões de

componente da Região 1, rotuladas 1A, 1B e 1C, podem ser identificadas, que são prognosticadas para preferencialmente influenciar Tyr20, Asn123 e Glu60 para as propriedades dos Resíduos-alvos respectivamente.

[000105] O termo "espaço da sequência" amplamente refere-se ao conjunto de todas as possíveis sequências de resíduos para um polímero de comprimento específico. Por exemplo, o espaço da sequência completo para uma proteína ou polipeptídeo de 100 resíduos de aminoácido em comprimento é o conjunto de todas as sequências que consistem em todas as possíveis variações dos 100 aminoácidos. Se forem consideradas mutações simples apenas, então o espaço da sequência para uma proteína ou polipeptídeo de 100 resíduos de aminoácido em comprimento é 100 vezes 19 (para o conjunto de 20 resíduos de aminoácido). Uma redução no espaço da sequência de uma proteína ou polipeptídeo de comprimento dado, portanto, reduzirá o número de possíveis sequências dos resíduos de aminoácido em um conjunto de proteína ou mutantes de polipeptídeo. No contexto da presente invenção, a identificação independente dos Resíduos de aminoácido-alvos e Resíduos de aminoácido Variantes produz dois conjuntos reduzidos diferentes de espaço da sequência a ser examinada. A interseção subsequente destes conjuntos causada quando os Resíduos de aminoácido-alvos são usados para identificar Resíduos de aminoácido Candidatos de Resíduos Variantes também reduz o espaço da sequência para consideração. Consequentemente, o número de resíduos da proteína e suas combinações que necessitam ser considerados para mutação é assim grandemente reduzido.

[000106] Como aqui usado, o termo "evolução direcionada" refere-se aos abordagens visados para modificar a função e/ou estrutura da(s) proteína(s) alvo(s). Em geral, evolução direcionada é um processo pelo qual as proteínas são "adaptadas" para atuarem em ambientes quí-

micos ou biológicos naturais ou artificiais diferentes ou existentes e/ou suscitar funções novas e/ou aumentar ou diminuir uma atividade dada, e/ou modular uma característica dada.

BREVE DESCRIÇÃO DAS FIGURAS

[000107] A presente invenção será agora descrita por via de exemplo com referência às figuras em anexo.

[000108] Fluxograma 1 fornece um fluxograma da estratégia de re-engenharia da Rubisco, destacando os papéis da química computacional e bioinformática fornecendo a informação mecanística, sequêncial, estrutural e filogenética que é integrada pelo procedimento de enxerto filogenético para produzir as predições *in silico* das proteínas Mutantes Candidatas. Esta integração fornece um meio pelo qual os Resíduos-alvos podem ser usados para selecionar os Resíduos Candidatos dos Resíduos Variantes, assim grandemente reduzindo o número de resíduos da proteína, e suas combinações que necessitam ser consideradas para mutação para melhoria da propriedade funcional. Resíduos Candidatos, como mostrados na coluna do meio, podem opcionalmente compreender ACRs e/ou DCRs. Estes procedimentos são mostrados em mais detalhes no Fluxograma 2. As etapas de predição são seguidas por triagem experimental e avaliação de melhoria da propriedade funcional. Os resultados podem ser realimentados no procedimento de predição para refinar as predições das proteínas Mutantes Candidatas, seguido por outros ciclos de triagem computacional e experimental e avaliação da melhoria da propriedade funcional, para otimizar a atividade da proteína. As caixas de linhas pontilhadas representam extensões opcionais para o procedimento de enxerto filogenético de núcleo que é mostrado em uma forma expandida no Fluxograma 3.

[000109] Fluxograma 2 fornece um fluxograma que detalha a estratégia de predição para o método de enxerto filogenético, mostrando a

integração da química computacional com a análise de bioinformática para produzir a lista classificada de proteínas Mutantes Candidatas. Este procedimento envolve etapas de selecionar Resíduos Variantes contra Resíduos-alvos para produzir Resíduos Candidatos (que opcionalmente podem compreender ACRs e/ou DCRs), agrupar e combinar os grupos de Resíduos Candidatos (e/ou Resíduos Candidatos Divergentes ou Resíduos Candidatos Alternativos), e classificar os grupos e grupos combinados para produzir uma lista classificada de proteínas Mutantes Candidatas para pré-triagem computacional opcional e triagem experimental.

[000110] Fluxograma 3 fornece um fluxograma que detalha uma extensão opcional do método para enxerto filogenético em que uma lista de Resíduos Candidatos é avaliada e montada em Mutantes Candidatos, e depois Resíduos Candidatos, Resíduos Candidatos Alternativos, Resíduos Candidatos Divergentes, Resíduos Covariantes e Resíduos Variantes Espécie-específicos podem também ser recrutados e agrupados, com base em sua influência nos Resíduos-alvos, e formados em Mutantes Candidatos refinados novos.

[000111] Figura 1 fornece uma ilustração das estruturas moleculares envolvidas no mecanismo de reação proposto para a conversão de RuBP para duas moléculas de 3-PGA no sítio ativo da Rubisco. As cinco etapas de enolização, carboxilação, hidratação, cisão da ligação de C_2-C_3 e protonação de C_2 são mostradas. Os numerais romanos correspondem às espécies intermediárias e do produto mostradas nas Figuras 3 e 4. Grupo R = -CH(OH)-CH_2-O-PO_3^{2}".

[000112] Figura 2 ilustra a estrutura molecular para o modelo de fragmento de 77 átomos do sítio ativo da Rubisco designado FM20 que foi usado para estudos químicos computacionais de QM *ab initio* da via de reação. Ele mostra as espécies de fragmento molecular que foram usadas para representar os resíduos de aminoácido LYS175, LYS177,

ASP203, GLU204, KCX201 (LYS201 carbamilado), HIS294, e LYS334, a molécula de água, a molécula de dióxido de carbono, e o fragmento de 4 carbonos da forma de enodiolato do substrato RuBP. Além disso, ele mostra os estados de carga das espécies de componente, e suas interações, a saber, os seis átomos coordenados para o átomo de Mg, ligações de hidrogênio e a interação de van der Waals entre CO_2 e C_2.

[000113] Figuras 3A a 3E fornecem uma série de estruturas moleculares calculadas usando abordagens de QM *ab initio* para o modelo de fragmento FM20 mostrado na Figura 2. Estas mostram geometrias de mínima local ao longo da via de reação de enodiolato para os produtos finais na reação de Rubisco carboxilase. Átomos de hidrogênio da molécula de água coordenada por Mg são destacados através de ciclos de preto. Distâncias relevantes para cada estrutura que demonstra as interações variáveis entre as espécies de componente à medida que a reação prosseguem são mostradas através de marcações. As marcações em algumas estruturas podem ser sobreadas. Marcação "d" refere-se a $R_{KCX201\text{-}H...O2}$, "i" refere-se a $R_{Ow...C3}$, "j" refere-se a $R_{H2O[Mg]\text{-}Hw1...O\text{-}GLU204}$ e "k" refere-se a $R_{H2O[Mg]...Hw1}$. Os numerais romanos correspondem ao intermediário, estado de transição e espécies de produto mostradas na Figura 4.

[000114] Figura 4 fornece um gráfico que ilustra a superfície de energia potencial para a carboxilação e reações subsequentes na via de carboxilase computada usando cálculos de química computacional de QM *ab initio* no modelo de fragmento com 77 átomos (FM20). Estágios diferentes do progresso da reação de carboxilase são distribuídos ao longo do eixo geométrico X. Os numerais romanos ao longo da via de reação designam os estados inicial (I) intermediário (III, V, VII) e do produto (IX) da reação de carboxilase mostrada na Figura 1, e seus estados de transição de conexão (II, IV, VI, VIII). As estruturas de todos estes estados são mostradas na Figura 3. Energias (em kcal/mol)

de todos os estados é mostrada com relação ao do estados de enediloato inicial I, e as energias do estado de transição com relação ao estado inicial e intermediário relevante são mostradas por Ea e uma seta.

[000115] Figuras 5A a 5D fornecem um alinhamento de sequência da sequências de aminoácido de LSU de Rubisco de organismos fotossintéticos que pertencem a treze diferentes filos abrangendo algas vermelhas, cianobactérias, glaucofitos e plantas (10 filos), usando simbologia de aminoácido de apenas uma letra. Estas sequências são também fornecidas como SEQ ID NOS: 1-13. Onde mais de uma sequência de Rubisco estava disponível em um filo, uma sequência de consenso de 50% foi usada para representar aquele filo. Para as sequências de consenso, o numeral entre parênteses representa o número de gêneros cujas sequências foram usadas para gerar a sequência de consenso. Os números de acesso da base de dados para todas as sequências usadas no alinhamento são dados na Tabela 1. A figura mostra que a sequência de LSU de Rubisco é altamente conservada, incluindo a ausência quase completa de intervalos com exceção das diferenças secundárias nos términos N e C, e códigos para uma cadeia de polipeptídeo longa de 475 resíduos (plantas e cianobactérias). O símbolo "~" denota intervalos de sequência. As letras maiúsculas de apenas uma letra denotam o alfabeto do resíduo de aminoácidos, enquanto as letras minúsculas e outros símbolos são mostrados nas posições nas sequências de consenso onde apenas o tipo de resíduo é conservado: "h", hidrofóbico (A, C, F, G, H, I, K, L, M, R, T, V, W, Y); "s", pequeno (A, C, D, G, N, P, S, T, V); "u", minúsculo (A, G, S); "a", aromático (F, H, W, Y); "c", carregado (D, E, H, K, R); "l", alifático (I, L, V); "p", polar (C, D, E, H, K, N, Q, R, S, T); "o", álcool (S, T); "t", semelhante a giro (A, C, D, E, G, H, K, N, Q, R, S, T); "-", negativamente carregado (D, E); e "+" positivamente carregado (H, K, R).

As sequências de consenso foram obtidas usando o servidor em http://coot.embl.de/ Alignment//consensus.html. O alinhamento foi corrigido por volta da posição de intervalo perto de CR T271 (posição de alinhamento 273) com base nas comparações estruturais entre as estruturas de raio X para espinafre (pdb 8ruc), *Synechococcus* (pdb lrbl) e *Galdieri partita* (pdb 1bwv) complexos com Mg_2^+ e 2CABP. As SEQ ID NOS para as estruturas são 17, 16 e 14, respectivamente, como mostradas na Tabela 1. Os números nos finais das linhas denotam o número da sequência. Os números uniformemente espaçados ao topo são marcadores de alinhamento, enquanto os números em negrito denotam o número da sequência para espinafre.

[000116] Figuras 6A e 6B fornecem sequências de consenso de 50% de LSUs de Rubisco de algas vermelhas (*rodófitas*; 9 espécies), cianobactérias (11 espécies) e plantas floríferas (*magnoliófitas*; 134 espécies). Estas sequências estão apresentadas na listagem de sequência como SEQ ID NOS: 2, 3 e 11, respectivamente, como mostradas na Tabela 1. Sombreamento cinza claro indica os 134 resíduos que são os mesmos em plantas floríferas e cianobactérias mas diferentes em algas vermelhas, isto é, os Resíduos Variantes. Intervalos na sequência são mostrados por "~". As definições de letras letras maiúscula e minúsculas e símbolos "-" e "+" são iguais às na Figura 5. Os números nos finais das linhas denotam o número da sequência. Os números uniformemente espaçados ao topo são marcadores de alinhamento, enquanto os números em negrito denotam o número da sequência para espinafre.

[000117] Figuras 7A e 7B fornecem sequências de consenso de 50% de LSUs de Rubisco de algas vermelhas (*rodófitas*; 9 espécies), cianobactérias (11 espécies) e plantas floríferas (*magnoliófitas*; 134 espécies), as mesmas sequências mostradas na Figura 6. Posições onde os resíduos de aminoácido são conservados entre todas as três

sequências de consenso são mostradas pelo símbolo de letra única para algas vermelhas (linha do topo) e pontos nas linhas para cianobactérias e plantas. Posições onde o resíduo de aminoácido é o mesmo para quaisquer cianobactérias e plantas como para algas vermelhas são mostradas pelo símbolo de apenas uma letra para algas vermelhas (linha do topo) e pontos nas linhas para cianobactérias ou plantas. Posições onde o resíduo de aminoácido é o mesmo em cianobactérias e plantas mas diferente em algas vermelhas, isto é, os 134 Resíduos Variantes mostrados na Figura 6, são mostradas pelo símbolo de apenas uma letra para algas vermelhas (linha do topo) e o símbolo de apenas uma letra para cianobactérias (segunda linha), enquanto a posição de resíduo tem um espaço em branco para plantas (terceira linha). Sombreamento reverso destaca aquelas posições de VR que foram selecionadas como Resíduos Candidatos para a Região 1, como mostradas na Tabela 2, e que foram agrupados em Mutantes Candidatos prognosticados, como mostrados nas Tabelas 3 e 4. Posições onde os resíduos de aminoácido são diferentes em todas as três sequências de consenso são mostradas pelos símbolos de apenas uma letra em todas as três linhas, com sombreamento cinza destacando aquelas posições selecionadas como Resíduos Candidatos Divergentes para a Região 1, como mostrados na Tabela 2, e que foram agrupados em Mutantes Candidatos prognosticados, como mostrados nas Tabelas 3 e 4. O símbolo "~" denota intervalos de sequência. Definições de letras maiúscula e minúsculas e símbolos "-" e "+" são iguais às na Figura 5. Os números nos finais das linhas denotam o número da sequência. O números uniformemente espaçados ao topo são marcadores de alinhamento, enquanto os números em negrito denotam o número da sequência para espinafre.

[000118] Figura 8 fornece uma vista lateral da estrutura do TIM barrel C-terminal (resíduos 151-475) de um polipeptídeo de LSU e o domínio

N-terminal (resíduos 1-150) de uma LSU adjacente que compreende a unidade abrigando um sítio ativo da Rubisco. Ele mostra as posições dos Resíduos-alvos, E60, N123 e Y20 para a Região 1, H294 para a Região 2 e K334 para a Região 3. Esta estrutura foi desenahda usando coordenadas de átomo da estrutura de raio X hexadecamérica de L_8S_8 completa da Rubisco de espinafre do tipo selvagem no complexo com Mg_2^+ e 2CABP (pdb 8ruc). 2CABP é mostrada como um modelo de CPK. Resíduos E60 e N123 ficam localizados na hélice αB e proximos da extremidade C-terminal da hélice αC, respectivamente, do domínio N-terminal. As cadeias laterais destes dois resíduos ficam localizadas em qualquer lado do grupo carboxilato de 2CABP. Hélices αB e αC têm interações hidrofóbicas com as β-fitas do domínio N-terminal. Y20 está localizado na estrutura de bobina simples próxima da extremidade N-terminal de βA. Tanto K334 como H294 estão localizados no domínio C-terminal contendo a maioria dos resíduos do primeiro-invólucro do sítio ativo. K334 está na alça 6, enquanto que H294 está em β5.

[000119] Figura 9 fornece uma vista da estrutura do domínio N-terminal do polipeptídeo de LSU de Rubisco (usando coordenadas da estrutura de raio X de espinafre 8ruc) com as cadeias laterais dos resíduos mutados (G23, R81, W25 e 151) para Mutante #6 (T23G, Y25W, D51I^{Gp}, K81R), e os Mutantes de componente #4 (T23G, K81R) e #1a (Y25W, D51I^{Gp}), modelados como nos modelos de bastão. Vide Tabela 3 para definições de Mutantes. Ela mostra as posições relativas destes resíduos na estrutura, e as interações entre os resíduos 25 (W25) e 51 (151), e entre os resíduos 23 (G23) e 81 (R81). As mutações adicionam uma interação hidrofóbica entre os resíduos W25 e 151, comparados com o Y25 e D51 do tipo selvagem, e removem uma interação de ligação de hidrogênio entre os resíduos 23 e 81 (T23 do tipo selvagem, K81 para G23 mutante, R81). Severamente ou em

combinação, estas duas mutações duplas são prognosticadas alterar Y20 para a orientação do resíduo. As posições dos Resíduos-alvos Y20, E60 e N123 são mostradas para referência como modelos de bola-e-bastão, enquanto o mímico intermediário da reação 2CABP é mostrado na forma de modelo de fio.

[000120] Figura 10 fornece uma vista da estrutura do domínio N-terminal do polipeptídeo de LSU de Rubisco (usando coordenadas da estrutura de raio X de espinafre 8ruc) com as cadeias laterais dos resíduos mutados (W25, 151, A54, A84, V87) para Mutante #7a (Y25W, D51 I^{Gp}, $G54A^{Gp}$, C84A, I87V), e os Mutantes de componente #1a (Y25W, $D51I^{Gp}$) e #5a ($G54A^{Gp}$, C84A, I87V), modelados em como modelos de bastão. A figura ilustra a formação de uma região hidrofóbica através de mutações nos sítios 54 (G -> A), 84 (C -> A) e 87 (I -> V). Adicionalmente, se G54 do tipo selvagem foi mutado para $S54^{Gm}$ em vez de $A54^{Gp}$ (Mutante #7b, #5b), então uma ligação de hidrogênio extra com a cadeia principal do resíduo 51 é prognosticada ser introduzida. Por sua vez, I51 têm uma interação hidrofóbica com o resíduo W25. Todas estas alterações são prognosticadas afetar o posicionamento do Resíduo-alvo Y20, mostrado na forma de bola-e-bastão. As posições dos outros Resíduos-alvos da Região 1 E60 e N123 são mostradas para referência como modelos de bola-e-bastão, enquanto o mímico intermediário da reação 2CABP é mostrado na forma de modelo de fio. Para referência, as posições das mutações para Mutante #4 (G23, R81) são também mostradas.

[000121] Figura 11 fornece uma vista do domínio N-terminal do polipeptídeo de LSU de Rubisco (usando coordenadas da estrutura de raio X de espinafre 8ruc) com as cadeias laterais dos resíduos mutados (V^{Gp} ou $I^{Gm}36$, L116, I121, L^{Gp} ou $I^{Gm}140$, G297, T300) para Mutante #10a (L36V, I116L, V121I, F140L, M297G, V300T) e outras variantes de Gp/Gm de Mutante #10 (#10b,c,d), e os Mutantes de compo-

nente #8 (V121I, M297G, V300T) e #9a (L36V, I116L, F 140L) e outras variantes de Gp/Gm de Mutante #9 (#9b,c,d), modelados como modelos de bastão. Resíduos A296 e V271 que não são mutados são também mostrados como modelos de bastão. Para facilitar entendimento das alterações nas interações prognosticadas para resultar das mutações, a inserção em caixas de RHS mostra os resíduos correspondentes em *Synechococcus* do tipo selvagem como modelo de bastão. A figura ilustra o rompimento prognosticado de uma interação hidrofóbica entre os resíduos V300, M297 e V121 no tipo selvagem para formar uma interação hidrofóbica nova entre A296, V271 e I121 no Mutante (#8 ou #10a). A predição é com base nas comparações estruturais entre as estruturas de raio X para complexos de espinafre (pdb 8ruc), *Synechococcus* (pdb 1rbl) e *Galdieri partita* (pdb 1bwv) com Mg_2^+ e 2CABP. Uma interação hidrofóbica nova entre os resíduos 36, 140 e 116 é também prognosticada pelas mutações L -> V, F -> L e I -> L. Severamente ou em combinação, estas interações hidrofóbicas novas mediadas por estes grupos de mutações são prognosticadas afetar a posição e orientação do Resíduo-alvo N123, que é mostrado no modelo de bola-e-bastão. As posições dos outros Resíduos-alvos da Região 1 E60 e Y20 são mostradas para referência como modelos de bola-e-bastão.

[000122] Figura 12 fornece uma vista estéreo do domínio N-terminal da LSU de Rubisco de espinafre (usando coordenadas da estrutura de raio X de espinafre 8ruc). As cadeias laterais de todos os resíduos N-terminais estão mostradas no modelo de estrutura de fio. Resíduos-alvos (Y20, E60 e N123) estão mostrados no modelo de bastão. Os análogos intermediários carboxilados 2CABP e Mg_2^+ estão mostrados como modelos de bola-e-bastão. A figura mostra uma representação 3-D completa de todas as cadeias laterais de resíduo do domínio N-terminal que são CRs selecionadas que são prognosticadas para influ-

enciar os TRs Y20, E60 e N123, como mostrado na Figura 13. Para referência, a orientação da estrutura é igual à na Figura 15.

[000123] Figura 13 fornece uma vista estéreo do domínio N-terminal da LSU de Rubisco de espinafre (usando coordenadas da estrutura de raio X de espinafre 8ruc), com os 20 Resíduos Candidatos e 6 Resíduos Candidatos Divergentes listados na Tabela 2 mostrados como modelos de bastão e os Resíduos Alvo rotulados N123, E60 e Y20. A figura mostra uma representação 3-D completa das cadeias laterais de CR e DCR no domínio N-terminal que são prognosticadas influenciar os TRs Y20, E60 e N123, como também um segmento pequeno do domínio C-terminal da LSU par mostrando os CRs M297 e V300 que são também prognosticados influenciar estes TRs. Resíduos rotulados T23 e K81 são aqueles que são mutados no Mutante #4 (T23G, K81R), enquanto os resíduos rotulados Y25 e E51 são adicionalmente mutados no Mutante #6 (T23G, Y25W, D51I^{Gp}, K81R).

[000124] Figuras 14A e 14B fornecem um alinhamento de sequência de aminoácido das sequências de consenso de 50% do domínio N-terminal das LSUs de Rubisco de plantas floríferas (*magnoliófitas*; 134 espécies), cianobactérias (11 espécies), e algas vermelhas (*rodófitas*; 9 espécies), como também sequências de várias plantas (espinafre, tabaco, arroz, soja e cana-de-açúcar), PCC6301 de *Synechococcus* sp. e espécies de algas vermelhas (*Galdieri partita* e *Griffithsia monolis*) de interesse especial. As sequências apresentadas nesta figura estão também expostas em SEQ ID NOS 2, 3 e 11, previamente usadas nas Figuras 5-7, e 14-21. Estas estão listadas com números de acesso de base de dados na Tabela 1. A sequência total é mostrada em apenas uma letra e forma de símbolo para o consenso das cianobactérias (vide Figura 5 para uma explanação dos símbolos), com resíduos de aminoácido em outras sequências mostradas por pontos se conservadas ou pelo formato de apenas uma letra/símbolo se não

conservadas. Os números nos finais das linhas denotam o número da sequência. Os números uniformemente espaçados ao topo são marcadores de alinhamento, enquanto os números em negrito denotam o número da sequência para espinafre. Faixas verticais sombreadas reversas pretas destacam os 17 Resíduos Candidatos atuais (isto é, aqueles resíduos comuns entre as sequências de consenso de 50% de plantas floríferas e cianobactérias mas diferentes em algas vermelhas) no domínio N-terminal: resíduos 18, 19, 23, 25, 51, 54, 59, 64, 68, 81, 84, 87, 88, 104, 114, 118, 121. Tiras verticais sombreadas cinza destacam os 6 resíduos Candidatos Divergentes atuais (isto é, aqueles resíduos que diferem em plantas verdes, cianobactérias e algas vermelhas): resíduos 36, 86, 116, 117, 138, 140.

[000125] Figura 15 fornece uma vista dos segmentos do domínio N-terminal e do domínio C-terminal da subunidade adjacente da LSU de Rubisco de espinafre (usando coordenadas da estrutura de raio X de espinafre 8ruc). Ao fundo o intermediário de reação β-ceto (1,5-bisfosfato de 2-carbóxi-3-cetoarabinitol (2C3KABP); III; vide Figura 1), Mg_2^+ e outros resíduos do domínio C-terminal da LSU par coordenados no primeiro-invólucro do sítio ativo (vide Figura 2) e sua disposição para os Resíduos-alvos Y20, E60 e N123. Y20, E60, N123 e o intermediário da reação β-ceto são mostrados como modelos de bastão grosso, enquanto os resíduos de sítio ativo são mostrados em modelo de bastão fino. A figura mostra para Sub-região IA no domínio N-terminal e Sub-região 1B no domínio N-terminal e segmentos do domínio C-terminal da subunidade adjacente. Sub-região IA inclui as posições de CR18 e novos CRs 19, 68 e 104 que são esperados melhorar o Mutante #4 (T23G/K81R) se mutado em combinações (#17-1A (#4 + K18I/T68V), #18-1A (#4 + P104E(D)), #19 (#4 + D19P)) e combinações também dos mesmos. Sub-região 1B inclui posições no novo Mutante Candidato #20 - 1B (V116L/L117T/V121I/I138M/F140L) con-

tendo DCRs 116, 117, 138 e 140, CR 121, e resíduo 135 totalmente conservado.

[000126] Figura 16A e 16B fornecem um gráfico com uma vista expandida das interações no Mutante Candidato #20-1B (V116L/L117T/V121I/ I138M/F140L) na Sub-região 1B. Painel A mostra para referência o domínio N-terminal completo, posições de TRs N123, E60 e Y20 e o centro do sítio ativo (2CABP, Mg_2^+), e hélice C e fita E. Painel B mostra como os resíduos de componente 116, 117, 120, 121 e 135, 138 e 140 localizados na hélice C e fita E interagem, e como alterações para estas interações são prognosticadas para influenciar perto de N123 do TR; outra vista é na Figura 17. 2CABP e Mg_2^+ são mostrados no modelo de CPK. N123, E60 e Y20 são mostrados no modelo de bastão grosso.

[000127] Figura 17 fornece um gráfico do domínio N-terminal da LSU de Rubisco e segmentos do domínio C-terminal da subunidade adjacente como um modelo de tira mostrando as posições relativa de todas as três Sub-regiões, 1A, 1B e 1C, e algumas das CRs e DCRs nestas Sub-regiões. Alguma das interações prognosticadas fundamentais usadas em designar os Mutantes Candidatos da Região 1 são mostradas por linhas conectando CRs e DCRs.

[000128] Figura 18 fornece um modelo estrutural usado em cálculos de QM/QM de ONIOM empreendidos com DFT como a camada alta e QM semiempírica (PM3) como a camada baixa. A camada de DFT (mostrada como modelo bola grande e de bastão) compreende 93 átomos incluindo fragmentos de todos os resíduos no primeiro invólucro de coordenação do sítio ativo e substrato: o átomo de magnésio (Mg_2^+), enodiolato de RuBP, GLU60, ASN123, LYS175, LYS177, LYS201 carbamilado, ASP203, GLU204, HIS294, LYS334. Esta camada do núcleo é rodeada por uma camada adicional de 711 átomos no PM3 (QM semiempírica) compreendendo resíduos de aminoácido

até ~12 Å de Mg_2^+. As coordenadas de partida foram tomadas da estrutura de raio X do complexo de Rubisco-2CABP de espinafre (pdb 8ruc). O tamanho do sistema inteiro são 804 átomos.

[000129] Figura 19 fornece um gráfico mostrando a estrutura do núcleo do sítio ativo da Rubisco e a divisão da estrutura que circunda os Resíduos-alvos em três Regiões espacialmente contíguas. Região 1 compreende resíduos de aminoácido que podem influenciar os TRs GLU60, ASN123 e TYR20 e podem compreender o domínio N-terminal e segmentos pequenos do domínio C-terminal da subunidade adjacente. Regiões 2 e 3 estão no domínio C-terminal e compreendem resíduos de aminoácido que podem influenciar os TRs HIS294 e LYS334, respectivamente.

[000130] Figura 20 fornece um gel que mostra a separação de Rubisco de L_2 de *R. rubrum* e Rubisco de L_8S_8 de tabaco mediante eletroforese de poliacrilamida não-desnaturante. Proteína de folha solúvel de tabaco do tipo selvagem, linhagens 1B Mutantes de tabaco #4 e Mutantes de tabaco #23 usadas para as medições de cinética foram separadas mediante eletroforese de poliacrilamida não-desnaturante e visualizadas por tingimento de Coomassie. A Rubisco de L_2 que está presente na linhagem de tabaco-*rubrum* de *ΔaadA* (tr) e na linhagem de transformante Mutante #4 de tabaco heteroplásmico #4 (isto é, produz tanto L_2 como Rubiscos de L_8S_8 mutados) não está presente nas outras linahgens Mutantes de tabaco ou transformante de tabaco do tipo selvagem. As proteínas marcadoras com tamanhos estão indicadas (m). A quantidade de proteína de folha solúvel carregada por rota é mostrada.

DESCRIÇÃO DETALHADA

[000131] A presente invenção fornece um método de gerar uma proteína com uma propriedade funcional melhorada. A invenção compreende um procedimento para estreitar o espaço da sequência para tes-

te mutacional convencional definindo mutantes com um ou, preferivelmente, várias mutações usando informação mecanística (computacional) e bioinformática e de base de dados (alterações de sequência filogenética-específica/ambiente-específica, dados cinéticos, estrutura 3-D e modelagem), em primeiro lugar. Este procedimento busca maximizar o uso de toda informação disponível, particularmente adaptações evolutivas parciais codificadas nas sequências de Rubisco, de forma que o processo de melhoria da proteína para a propriedade funcional requerida inicia em um nível funcional mais próximo do nível requerido de eficiência funcional.

[000132] Uma primeira etapa do método compreende o processo de identificar Resíduos de aminoácido-alvos em uma primeira proteína. Como descrito em mais detalhe abaixo, o processo de identificar Resíduos de aminoácido-alvos pode compreender cálculos de QM de fragmento de sítio ativo (etapa (i) usando, por exemplo, abordagens de DFT), e cálculos de QM/QM híbrida ou de QM/MM (etapa (ii) usando, por exemplo, abordagens de ONIOM), junto com uso de dados empíricos (por exemplo, dados cinéticos e estruturas de cristal de raio X). Simulações da dinâmica molecular (MD) (etapa (iii)), e cálculos usando combinações de QM, QM/QM ou QM/MM, e abordagens de MD (etapas (iv) e (v)) podem ser usados, respectivamente, para avaliar a estabilidade das proteínas Mutantes Candidatas prognosticadas, ou para entendimento mais detalhado dos papéis dos Resíduos-alvos na reação enzímica. Estes Resíduos de aminoácido-alvos usualmente serão resíduos de base para o funcionamento apropriado da proteína com respeito à propriedade funcional sob investigação. É uma característica preferida da invenção que a propriedade funcional melhorada seja alcançada modificando as propriedades químicas destes resíduos (por mutação de outros resíduos) para melhorar, por exemplo, sua atividade cinética.

CONVENÇÃO PARA NUMERAÇÃO DE RESÍDUOS DE AMINÁCIDO DE RUBISCO

[000133] Ao longo deste relatório descritivo, ao identificar um resíduo de uma LSU de Rubisco por número, a numeração do resíduo foi com base na numeração do resíduo da sequência de aminoácido de LSU de Rubisco de espinafre (SEQ ID NO: 17). Esta convenção de numeração foi usada para todos os resíduos identificados em química computacional, em estruturas de proteína e em mutações. Isto não cria ambiguidade como estes números podem ser mapeados para números de sequência em alinhamentos e de comparações estruturais, e consequentemente um número de resíduo de espinafre dado pode ser mapeado com confiança total para um número de resíduo de cianobactérias ou algas vermelhas estruturalmente equivalente.

LISTAGENS DE SEQUÊNCIA

[000134] Tabela 1 fornece um sumário das sequências de aminoácido da Rubisco de ocorrência natural e de consenso de 50% debatidas aqui e que são fornecidas na listagem de sequência legível por computador, com as SEQ ID NOS como mostradas na coluna "SEQ ID". Onde sequências múltiplas foram usadas para produzir uma sequência de consenso de 50%, o número total de sequências envolvidas na criação da sequência de consenso é listado entre parênteses na coluna "Descrição". Os números de acesso das bases de dados fornecem identificadores únicos para cada uma das sequências que foram usadas, incluindo aquelas sequências que foram consideradas na criação das sequências de consenso. Os números das figuras em que as sequências são usadas nos alinhamentos são também dados na coluna "Figura Nos".

[000135] TABELA 1: Números de acesso das bases de dados para Sequências de LSU de Rubisco usadas nas Figuras 5, 6, 7 e 14. Os números ID da sequência correspondem àqueles na listagem de sequência legível por computador.

SEQ ID NO:	Descrição	Nos da Figura	Número(s) de acesso das bases de dados)			
1	Protista glaucófitas	5	P24312			
2	Protista rodófitas (9)	5,6,7,14	BAA75676 ABU53651 BAE78417	AAB17222 CAB58236	BAA75796 AAD04746	AAR13681 BAE78409
3	Eubacteria cianobactéraias (11)	5,6,7,14	P008795 199140 4296100	400856410 P27568 403238160	4797620 403124030 P00880	402041420 Q8DIS5
4	Plantae antocerotófitas	5	Q31795			
5	Plantae briófitas (28)	5	Q95G53 Q76GQ0 Q76H L3 Q94N80 Q8HW62 Q5TM99 Q5TM97	Q76GQ2 Q50L57 Q5TM93 Q5TMA1 Q8SN97 Q5TMA5 Q9GGM2	Q5TM96 Q5TMB1 Q9BB41 Q9TM58 Q95G62 Q9GIF5 Q5TM95	Q75W61 Q75W60 Q95G63 Q75VP7 Q9TM63 Q5TMA3 NP_904194
6	Plantae carófitas(4)	5	Q8SN66	P48716	Q32RY7	Q32RQ1
7	Plantae clorófitas(4)	5	NP_958405	BAE48225	AAD00447	BAC06367
8	Plantae coniferófitas	5	P41621			
9	Plantae equisetófitas	5	P48702			
10	Plantae gnetófitas	5	Q9THI3			
11	Plantae magnoliófitas (134)	5,6,7,14	Q3T5C7 Q31857 Q3V6M3 Q5I221 Q6R615 P48690 Q9XQE3 Q5EKL6 CAA57001 Q9XQB9 Q9XQE7 Q3L237 Q33449 Q9BBC7 Q9GGC1 Q95F23 Q32488 Q75VD6 Q9GHT6 Q42916 Q32685 Q9GHS0 Q37257 Q95A48 Q9XK53 Q5MB28	Q5EKM0 Q9MRW9 Q31670 Q06022 Q3T5C1 Q75VD8 Q9XPK2 063123 Q42674 Q6R613 Q5EKM4 Q5EKM2 Q9ZT30 P19161 Q42828 Q3T5G1 Q5EKL5 Q9BBU1 Q8WIB0 098611 NP_054507 Q68RZ8 Q36849 Q32916 Q9XQA7 Q5EKL2	Q95EI0 Q3V6P6 063085 Q37167 P48688 Q42664 Q8WJD8 Q8SLM3 P92255 Q32072 Q8WLJ3 P92287 Q32188 P48703 Q4VWN7 Q95BC5 AAK72524 Q32518 Q32622 Q5EKL8 Q5C9P7 Q3T5F6 Q95F15 P04717 Q32820 Q8M962	Q06021 Q9GDM8 Q3T5G3 AAF78948 Q06023 Q7YKF9 P48693 Q7YKF8 O98530 Q3T575 Q95F13 Q37319 Q9SB16 CAB08877 Q5XLF7 Q95F12 Q01873 $Q95EH_6$ Q9MVF1 Q3T5E7 BAA00147 Q7YL87 Q95F24 Q8WGU4 Q8ME88 Q75VD3
			O98612	Q8WKT8	Q6R614	Q9MTS7
			Q8WIA8	Q95F20	AAX44985	O62943
			Q8WIC4	P48715	ABB90049	Q9XQ93
			Q9GHN0	Q3T5E2	Q33064	Q37281
			Q8WIC3	Q9XPK3	Q6R617	Q8LUX7
			Q8WKR2	AAP92166	Q6USP5	Q3T5E4
			P28459	Q95F10	P92364	CAA60294
			Q75VD7	NP_054944		

12	Plantae Pinófitas(2)	5	P26961	P26962		
13	Plantae Pteridófitas(2)	5	Q85WR7	Q33015		
14	Galdieria partita	14	BAA75796			
15	Griffithsia monolis	14	ABU53651			
16	PCC6301 de Synechococcus elongatus	14	P00880			
17	Espinafre (Spinacia oleracea)	14	NP_054944			
18	Tabaco (Nicotiana tabacum)	14	NP_054507			
19	Arroz (Oriza sativa)	14	BAA00147			
20	Soja (Glycine max)	14	YP_538747			
21	Cana-de-açúcar (Saccarum officinarum)	14	BAD27301			

IDENTIFICAÇÃO DOS RESÍDUOS-ALVOS - MECANISMO COMPUTACIONAL

[000136] Etapas (i) e (ii), daqui por diante com relação à identificação dos Resíduos de aminoácido-alvos, foram executadas usando o pacote de programa GAUSSIAN, para cálculos de QM e de ONIOM *ab initio* (Frisch et al, 2004), mas há vários outros programas particulares ou livre-para-usar disponíveis que poderiam ser usados alternativamente. Etapa (iii) usa o programa AMBER em geral disponível (Case et al, 2006) para executar as simulações de MD da proteína; esta capacidade está também disponível em muitos outros programas. Etapas (iv) e (v) empregam teoria, protocolos e programas publicados para simulações de mecanismo da enzima (Gready et al, 2006); os abordagens de simulação de MD de QM/MM semiempírica de núcleo (Cummins e Gready, 1997, 1998, 1999, 2003, 2005; Cummins et al, 2007) são implementados no programa MOPS (Cummins, 1996).

(i) CÁLCULOS DE QM DE COMPLEXOS DE FRAGMENTO DE SÍTIO ATIVO

[000137] A descrição a seguir refere-se aos cálculos em relação à LSU de Rubsico em que os resíduos de sítio ativo são totalmente conservados entre as espécies. Estes cálculos usam um método de QM de nível alto *ab initio* (B3LYP/6-31G(d,p)) para definir os energéticos e

estruturas das espécies de reação (substrato, estado de transição (TS), intermediário, e complexos de produto) no mecanismo de reação da Rubisco de multi-etapas, como mostrado na Figura 1. Computações foram executadas para as etapas de reação a partir da reação de adição de gás, usando o conjunto de programas Gaussian 03 (Frisch et al, 2004) e com coordenadas de início tiradas da estrutura de raio X do complexo de Rubisco-2CABP de espinafre (pdb 8ruc). O modelo de fragmento de sítio ativo, FM20 como mostrado na Figura 2, era grande o bastante para conter todos os resíduos e moléculas de água que interagem imediatamente com o centro de reação e o átomo de Mg coordenado no sítio ativo, e permitir definição de seus papéis nas etapas de reação diferentes, incluindo as etapas de adição de gás fundamentais (CO_2 ou O_2). As estruturas das espécies de reação e a via de energia da reação para as etapas de carboxilação e de reação subsequentes estão mostradas nas Figuras 3 e 4, respectivamente.

(ii) CÁLCULOS DE QM/QM HÍBRIDA DE ONIOM E DE QM/MM:

[000138] Estes cálculos definem as perturbações para a energética e estruturas das espécies da via de reação, e principalmente focalizou na etapa de adição de gás do próximo mais próximo dos vizinhos e além dos resíduos de sítio ativo. Isto foi feito usando abordagens que usam um modelo de QM alto, mas computacionalmente caro, *ab initio* para o núcleo do sistema (isto é, como em (i)) e um QM menos cara (QM semiempírica) ou Modelo de MM para uma região estendida.

[000139] Os cálculos (QM/QM e QM/MM) foram executados em vários estágios usando o módulo de ONIOM em GAUSSIAN 03. Os cálculos de QM/QM de ONIOM usaram um modelo de uma camada de núcleo de QM de alto nível *ab initio* de 93 átomos. Para estudo do ponto de partida na etapa de adição de gás, a camada de núcleo de QM compreende o átomo de magnésio (Mg_2^+), enodiolato de RuBP (para computar a estrutura e energias de espécies de reação subsequentes

as espécies químicas derivadas de RuBP correspondentes foram usadas), GLU60, ASN123, LYS175, LYS177, LYS201 carbamilado, ASP203, GLU204, HIS294 e LYS334. A camada de núcleo é rodeada por uns 711 outros átomos na camada externa computada no nível de PM3 (QM semiempírica) compreendendo resíduos de aminoácido até ~12 Å do átomo de magnésio. As coordenadas de partida foram tiradas da estrutura de raio X do complexo de Rubisco-2CABP de espinafre (pdb 8ruc). Este modelo está ilustrado na Figura 18. O papel destes cálculos são comparar os efeitos dos resíduos enxertados na redondeza do sítio ativo com aqueles do tipo selvagem na energética e estruturas das espécies da via de reação, principalmente focalizado na etapa de adição de gás. Estes cálculos permitiram as estruturas e energias dos fragmentos de sítio ativo de 93 átomos das espécies de reação, análogos aos para o modelo de QM de FM20 mostrado nas Figuras 3 e 4, ser reotimizado no ambiente dos resíduos de enzima circunvizinhos (a camada externa de 711 átomos), e, desse modo, permitiu perturbações para o perfil de energia da reação devido aos resíduos enxertados a serem estimados. Estes cálculos podem, desse modo, fornecer valiosos entendimentos nas perturbações para o mecanismo básico de, neste caso, Rubisco de espinafre do tipo selvagem a ser determinada, e, em particular, os detalhes da estrutura de CO_2 ligado e o grupo carboxilato nascente na etapa de adição de gás. Por este meio, a magnitude e direção das perturbações eletrostáticas das interações dos Resíduos-alvos com espécies de reação devido aos resíduos enxertados em Mutantes Candidatos podem ser calculadas. Esta informação pode ser usada para pré-triar Mutantes Candidatos para teste experimental e/ou usados na interpretação de resultados de teste experimentais, como mostrado no Fluxograma 1, coluna à esquerda.

(iii) SIMULAÇÕES DA DINÂMICA MOLECULAR (MD) DOS COM-

PLEXOS DE PROTEÍNA

[000140] Estas simulações avaliaram se a estrutura da proteína de Mutantes Candidatos de Rubisco enxertados poderia acomodar os resíduos alterados, isto é, se a estrutura de proteína mutante era conformacionalmente estável ou tendia a desfazer-se. Simulações de MD são particularmente úteis para vários mutantes enxertados, e fornece um teste de triagem de estabilidade global para complementar os testes eletrônicos no mecanismo químico de (ii).

[000141] Estes cálculos foram executados com o pacote de programa AMBER8 ou AMBER9 (Case et al., 2006), mas outros pacotes de simulação de MD de proteína (por exemplo, GROMACS) poderiam ser usados para obter resultados similares.

(iv) MÚLTIPLOS CÁLCULOS DE QM/QM HÍBRIDA OU DE QMMM DE ONIOM DE DIFERENTES ESTADOS CONFORMACIONAIS AMOSTRADOS DOS COMPLEXOS

[000142] Neste método, uma série de cálculos é empreendida usando coordenadas para os complexos de proteína (por exemplo, para etapas de reação diferentes) tiradas de tomadas instantâneas das trajetórias de simulações de MD de QM/MM, como descritas por Gready et al. (2006). Estes cálculos permitiram um exame mais detalhado das características da via catalítica, a saber os efeitos da flexibilidade conformacional da proteína nas geometrias do complexo de enzima e as energias de ativação e reação.

(v) GERAÇÃO DAS SUPERFÍCIES DE ENERGIA LIVRE DE REAÇÃO TOTAL PARA AS REAÇÕES DE ADIÇÃO DE GÁS

[000143] Um conjunto estatístico completo (média conformacional) dos estados de enzima no curso completo de uma etapa de reação pode ser gerado de simulações de MD de QM/MM semiempírica para uma grade de pontos definidos pelas coordenadas de reação (uma hipersurfície de energia livre). Uma superfície de energia livre mais

precisa pode depois ser gerada no nível de QM *ab initio* por cálculos de QM/MM de ONIOM usando configurações múltiplas, por exemplo, até 120, pontos testados na hipersurperfície de reação de QM/MM semiempírica (Gready et al, 2006; Cummins et al, 2007). Estas superfícies enzímicas de energia livre fornecem energias livres de reação e ativação que podem ser comparadas diretamente com os dados experimentais, tais como constantes cinéticas experimentalmente medidas, e também podem ser usadas para calcular as diferenças nas energias livres de reação e ativação entre tipo selvagem e mutantes.

(vi) DEFINIÇÃO DO MECANISMO DE REAÇÃO E RESÍDUOS-ALVOS

[000144] Com base nos resultados destas computações, os inventores puderam deduzir um mecanismo para a sequência inteira de reações na catálise de carboxilase, e definir papéis precisos para os resíduos de sítio ativo, isoladamente e em combinação. Dos cálculos de fragmento de QM, identificou-se um par de resíduos de aminoácido fundamentais, um atuando como uma base e o outro atuando como um ácido, para cada etapa de reação. Em particular, o par HIS294 e LYS334 foram identificados para a etapa de adição de gás.

[000145] O ponto de partida é o complexo de Rubisco com a forma de enodiolato de RuBP ligado à reação de Rubisco carboxilase, para o sítio ativo e a molécula de CO_2 retida em uma distância de interação de van der Waals para o carbono C_2 do enodiolato. Este estado é representado como I nas Figuras 1, 3 e 4. A reação prossegue através de um estado de transição caracterizado pela formação de uma ligação covalente parcial entre o átomo de carbono CO_2 e o carbono C2 do enodiolato, acompanhada por ligações duplas parciais entre os átomos de carbono C2 e C3 e entre os átomos C3 e O3 do enodiolato. Este estado é representado como II nas Figuras 3 e 4. A etapa de reação termina com a formação completa de uma ligação covalente entre

a molécula de gás e o carbono C2 do intermediário β-ceto (2C3KABP). Este estado é representado como III nas Figuras 1, 3 e 4.

[000146] Na etapa de adição de gás, HIS294 atua como uma base para completamente remover um próton do átomo O3; isto transfere uma carga negativa parcial ao carbono C2 e o possibilita formar uma ligação covalente com o átomo de carbono CO_2, e também transfere a carga negativa ao grupo carboxilato nascente. LYS334 que está positivamente carregado ajuda a estabilizar esta carga negativa que desenvolve no grupo carboxilato nascente. Estas características podem ser vistas nas estruturas detalhadas para I - III na Figura 3. Consequentemente, a basicidade de HIS294 e a acidez (carga) de LYS334 são cruciais para a etapa de adição de gás.

[000147] Modificando as propriedades destes dois resíduos, por exemplo, estericamente alterando para a orientação/distância de suas interações com o substrato de enodiolato ou intermediário β-ceto ou eletronicamente alterando a carga nos átomos que interagem com o substrato de enodiolato ou intermediário β-ceto pode afetar o energética da etapa de adição de gás. HIS294 e LYS334 são, desse modo, identificados como "Resíduos-alvos", amplamente definidos ser resíduos prognosticados por terem um efeito significativo no mecanismo de reação e energética. HIS294 e LYS334 estão no domínio C-terminal, são espacialmente separados, e afetam partes diferentes do substrato de enodiolato ou intermediário β-ceto. Consequentemente, os aminoácidos que podem afetar suas propriedades foram classificados em regiões diferentes; Região 2 para His294 e Região 3 para Lys334, como mostradas na Figura 19.

[000148] Embora o resíduo ASN123 não tenha sido incluído no modelo de fragmento de sítio ativo de FM20 para os cálculos de QM, exame da estruturas de cristal e os cálculos de QM/QM preliminares sugeriram que ele também esteja envolvido na estabilização da carga

no grupo carboxilato nascente acrescentado a C2. Além disso, exame de estruturas de cristal mostraram que os resíduos E60 e Y20 são posicionados para alterar a carga diretamente em LYS334 (isto é, a carga/orientação de LYS334 podem ser alterados manipulando E60 e Y20), e o grupo C_2-carboxilato do intermediário, 2C3KABP. Desse modo, E60, Y20 e N123 foram também identificados como Resíduos-alvos. Estes três resíduos estão no domínio N-terminal da LSU e, desse modo, aminoácidos que podem afetar suas propriedades foram classificados como pertencendo a uma região diferente (Região 1) daquela de HIS294 e LYS334 (Figura 19). Estes três resíduos são estritamente conservados em todas LSUs de Rubisco cataliticamente ativas e prognosticados atuarem de uma maneira combinada na etapa de adição de gás. Isto pode ser medido mediante inspeção das Figuras 8, 9, 10, 12 e 15 que fornecem vistas diferentes da disposição relativa de ASN123, GLU60 e Tyr20 com respeito ao grupo carboxilato do análogo do intermediário β-ceto (2C3KABP). Estas figuras mostram que ASN123, GLU60 e Tyr20 estao presos a três regiões de estrutura secundária separada do domínio N-terminal e que suas cadeias laterais estendem-se no sítio ativo em uma constelação tripartida.

[000149] Em resumo, o método acima compreende um conjunto total de abordagens computacionais para investigar problemas mecanísticos, enérgicos e de estabilidade em níveis globais ou mais detalhados para as etapas de carboxilação e de oxigenação para tipo selvagem e qualquer Mutante Candidato de Rubisco prognosticados. Desta maneira, foi possível identificar um ou mais Resíduos-alvos para atuar como o foco para o enxerto filogenético.

COMPARAÇÕES DE PROTEÍNA - ENXERTO FILOGENÉTICO

[000150] Em sua forma mais vasta, o método descrito aqui também compreende a comparação de pelo menos uma segunda proteína com pelo menos a primeira proteína. A segunda proteína pode originar-se

da mesma ramificação filogenética ou uma diferente que a primeira proteína. O processo de comparação vincula-se à identificação de pelo menos um Resíduo de aminoácido Variante entre a primeira proteína e a segunda proteína. Uma pluralidade de Resíduos Variantes da segunda proteína atua como um fundo geral de identidades de resíduos de aminoácidos específicos diferentes que podem ser "enxertados" sobre a primeira proteína em uma tentativa para melhorar a propriedade funcional da primeira proteína mediada pelos Resíduos-alvos.

[000151] Tomando Rubisco como um exemplo, alterações específicas de ramificação filogenética na sequência de aminácido de Rubisco, tal como alterações nas Rubiscos dos grupos filogenéticos de linhagens evolutivas diferentes ou nas Rubiscos que expressam alterações ambiente-específicas, representam possíveis otimizações parciais da eficiência catalítica das Rubiscos. Uma estratégia, denominada "enxerto filogenético", foi desenvolvida para identificar os resíduos fundamentais que representam estas soluções evolutivas parciais e para seletivamente "transplantar" estes resíduos em uma Rubisco do hospedeiro, tal como uma Rubisco de *Synechococcus* sp., alterando os resíduos hospedeiros específicos para aqueles do grupo doador de Rubisco ou doador de Rubiscos com uma ou mais características cinéticas melhoradas (ou preferidas), com uma vista para produzir uma Rubisco do hospedeira com estas características cinéticas melhoradas.

[000152] Soluções evolutivas parciais descritas acima foram identificadas por um procedimento de combinar os resultados dos estudos computacionais (os Resíduos-alvos), como mostrados no Fluxograma 1, coluna à esquerda, com aqueles da análise filogenética (os Resíduos Variantes), como mostrados no Fluxograma 1, coluna à direita, para selecionar os Resíduos Candidatos, como mostrados no Fluxograma 1, coluna do meio. Estas soluções são distribuídas entre os Re-

síduos Candidatos em alterações de sequência conservadas características (consensos) da LSU entre ramificações filogenéticas diferentes das Rubiscos, ou em alterações entre Rubiscos da mesma ramificação que são melhores adaptadas a ambientes específicos, por exemplo, seco/úmido ou quente/frio.

[000153] A integração dos resultados dos estudos computacionais com aqueles da análise filogenética para identificar um subconjunto específico de Resíduos Variantes (isto é, os Resíduos Candidatos) permite diferenciação entre os resíduos, que pode afetar uma propriedade funcional, por exemplo, a eficiência da etapa de adição de gás, de outras alterações de sequência conservadas características (consensos) entre as ramificações filogenéticas, que pode representar por exemplo, desvio filogenético neutro ou um papel fisiológico ramificação-específico. Tomando o exemplo de uma enzima de Rubisco, um papel fisiológico ramificação-específico pode incluir enovelamento e montagem da proteína, incluindo interações com a subunidade pequena, ou estabilidade da proteína.

(i) IDENTIFICAÇÃO DE RESÍDUOS VARIANTES POR ANÁLISES FILOGENÉTICAS

[000154] O uso combinado dos mecanismos computacionalmente deduzidos para identificar Resíduos-alvos com conservação de sequência e informação filogenética para identificar os Resíduos Variantes é ilustrado pelo debate a seguir de fatores de especificidade de Rubisco. Os fatores de especificidade muito alta de Rubiscos algais vermelhas podem ser atribuídos aos resíduos que estão em comum entre as Rubiscos das cianobactérias e de plantas floríferas, mas diferem em Rubiscos algais vermelhas. Tais resíduos são definidos aqui como "Resíduos Variantes". Se Resíduos Variantes simples ou uma pluralidade de Resíduos Variantes que atuam como fatores determinantes da especificidade nas Rubiscos algais vermelhas forem identifi-

cados e seletivamente incorporados em Rubiscos de plantas floríferas/de cianobactérias, então uma Rubisco que é fisiologicamente ativa no organismo hospedeiro pode ser produzida tendo especificidade mais alta para CO_2 que a enzima nativa. Resíduos do primeiro invólucro, isto é, aqueles resíduos que coordenam diretamente para o centro da reação (Glu60, Asn123, Lys175, LYS177, KCX201, Asp203, Glu204, His294 e Lys334) são totalmente conservados entre as Rubiscos. Esta conservação é ilustrada na Figura 5 que mostra um alinhamento das sequências de LSU de Rubisco de organismos fotossintéticos que pertencem a treze filos diferentes abrangendo algas vermelhas, cianobactérias, glaucófitas e plantas (10 filos). Onde mais de uma sequência de Rubisco estava disponível em um filo, uma sequência de consenso de 50% foi usada para representar aquele filo. As sequências de consenso foram obtidas usando o servidor em http://coot.embl. de/Alignment//consensus.html. Figura 5 também mostra que a sequência de LSU de Rubisco de 475 resíduos (plantas e cianobactérias) é em geral altamente conservada, incluindo a ausência quase completa de intervalos com exceção de diferenças secundárias nos términos N e C.

[000155] Porém, resíduos no segundo e invólucros subsequentes que circundam o centro da reação mostram variação entre as ramificações principais da Rubisco de plantas floríferas, algas vermelhas e cianobactérias. Algas vermelhas mostram a maior especificidade para CO_2, como identificadas pela razão de CO_2/O_2 de ~160 comparada com ~80 para plantas verdes e ~40 para cianobactérias. A variação de sequência entre as plantas floríferas, algas vermelhas e cianobactérias é ilustrada mais claramente no alinhamento na Figura 6 que compreende apenas as sequências de consenso de 50% de LSUs de Rubisco de algas vermelhas (*rodófitas*; 9 espécies), cianobactérias (11 espécies) e plantas floríferas (*magnoliófitas*; 134 espécies), já mostradas na

Figura 5. Figura 6 mostra que há 134 resíduos que são as mesmas em plantas floríferas e cianobactérias mas diferentes em algas vermelhas, isto é, os Resíduos Variantes. Os números de acesso da base de dados para cada uma das sequências usadas nos alinhamentos nas Figuras 5 e 6 são dados na Tabela 1, juntamente com as SEQ ID NOS para as listagens de sequência legíveis por computador.

[000156] Neste exemplo, foram identificados 134 Resíduos Variantes da LSU de Rubisco, entre os quais são distribuídos os resíduos que são responsáveis pela solução evolutiva parcial para especificidade aumentada que é exibida por Rubiscos algais vermelhas. Isto é mostrado como resíduos sombreados cinzas no alinhamento na Figura 6. Como os fatores determinantes de especificidade podem ser codificados por combinações de vários Resíduos Variantes, muitos milhares de tais combinações são possíveis. Consequentemente a fim de ser de qualquer uso prático, o subconjunto compreendendo os determinantes de especificidade necessita ser selecionado da lista de Resíduos Variantes.

(ii)(a) IDENTIFICAÇÃO DOS RESÍDUOS CANDIDATOS

[000157] Usando os abordagens descritos abaixo, Resíduos Variantes específicos foram identificados os quais têm o potencial para afetar a etapa de adição de gás da reação catalisada por Rubisco, e estes foram denominados "Resíduos Candidatos". Isto permitiu alterações conservadas entre as ramificações ou sub-ramificações filogenéticas/sub-espécies, que pode representar desvio filogenético neutro ou que pode ter um papel fisiológico ramificação-específico, tal como na estabilidade, enovelamento ou conjunto da LSU de Rubsico, a ser desconsiderado.

[000158] Muitos dos Resíduos Variantes podem não contribuir para a propriedade melhorada da Rubisco e por conseguinte podem não fazer parte da solução evolutiva, neste exemplo, para especificidade de

CO_2 aumentada, mas do contrário são mutações silenciosas ou mutações relevantes para outras propriedades da enzima, tais como o enovelamento e montagem da proteína, ou sua estabilidade, na célula. Para identificar os Resíduos Variantes mais prováveis de fazer parte da solução evolutiva para a propriedade melhorada, neste exemplo especificidade aumentada de CO_2, os entendimentos mecanísticos obtidos dos cálculos de QM foram empregados para selecionar Resíduos Candidatos da pluralidade de Resíduos Variantes. Este procedimento foi com base na hipótese que Resíduos Variantes que podem influenciar a funcionalidade dos Resíduos-alvos identificados pela etapa de química computacional como envolvidos na etapa de adição de gás em Rubisco, formam uma parte da solução evolutiva. Este foi o critério primário usado para selecionar os Resíduos Variantes para se obter um subconjunto de resíduos aqui chamado "Resíduos Candidatos". Em geral, o processo de seleção foi com base em avaliar a proximidade espacial dos Resíduos variantes para os Resíduos-alvos e estimar e classificar sua habilidade para influenciar as eletrostáticas e orientação dos Resíduos-alvos. Seleção pode utilizar triagem visual de estruturas cristalográficas usando uma modelagem molecular e pacote de programa de visualização tal como Accelrys Discovery Studio v2.0 (Accelrys Software Inc., San Diego, CA, 2007), embora outros pacotes de modelagem similares pudessem ser usados. Conceitos químicos padrão para interações intermoleculares, tais como preamento eletrostático de carga-carga, distâncias típicas de van der Waals e de ligação de hidrogênio, e modelos de enchimento espacial podem ser usados para cadeias laterais de aminoácido em uma varredura inicial de resíduos para seleção. O procedimento pode também ser sistematizado, por exemplo, mapeando todas as interações eletrostáticas e hidrofóbicas de átomo-para-átomo de cada um dos Resíduos Variantes com todos os outros resíduos de aminoácido que estão dentro de 3-5 Å em

distância e excluindo aquelas interações que são equivalentes nas sequências da primeira e segunda proteína.

[000159] Exemplos de tais interações equivalentes incluem interações de cadeia principal-cadeia principal que são em geral, mas nem sempre, inalteradas pela mutação. Interações foram também consideradas equivalentes, por exemplo, se uma interação hidrofóbica na sequência de uma proteína envolvesse um carbono α ou β alifático da cadeia lateral de um Resíduo Variante com um átomo de um Resíduo não-variante, e na sequência da segunda proteína o aminoácido Resíduo Variante, embora diferente daquele na sequência da outra proteína, também teve um carbono α ou β alifático na cadeia lateral interagindo com o mesmo Resíduo não-variante como na primeira sequência de proteína.

[000160] Interações dos grupos metila nas cadeias laterais de aminoácido, tais como aquelas em valina, leucina e isoleucina foram consideradas equivalentes se apenas o grupo metila particular estivesse envolvido na interação com um Resíduo de não-variante. Ligações de hidrogênio formadas pelos grupos carboxilato de aspartato e resíduos de glutamato foram também consideradas equivalentes se as distâncias de ligação de hidrogênio correspondentes fossem similares.

[000161] Após triar os Resíduos Variantes para diferenças em padrões de interação entre as sequências da primeira e segunda proteína, apenas aqueles Resíduos Variantes que tiveram o potencial para afetar um Resíduo-alvo identificado através de padrões de interação alterados foram retidos. O potencial de um Resíduo Variante para afetar um Resíduo-alvo foi reconhecido pela interação de um Resíduo Variante com o Resíduo-alvo ou com resíduos de aminoácido adjacentes para Resíduos-alvos ou com resíduos de aminoácido na unidade estrutural secundária abrigando o Resíduo-alvo. Até mesmo Resíduos Variantes que fazem parte das alças, voltas ou fitas não-estruturados,

mas que são conectados às unidades estruturais secundárias abrigando o Resíduo-alvo, têm o potencial para alterar a orientação de um Resíduo-alvo ajudando no reposicionamento das unidades estruturais secundárias através de interações alteradas. Os Resíduos Variantes selecionados que tiveram uma ou mais interações variantes e o potencial para influenciar Resíduos-alvos constituíram o conjunto dos "Resíduos Candidatos". Os 20 Resíduos Candidatos identificados como capazes de influenciar os Resíduos-alvos na Região 1, isto é, ASN123, Glu60 e TYR20, são identificados através de sombreamento reversos no alinhamento na Figura 7, e são mostrados graficamente na Figura 13. Eles estão também listados na Tabela 2.

(ii)(b) IDENTIFICAÇÃO DE RESÍDUOS CANDIDATOS ALTERNATIVOS E RESÍDUOS CANDIDATOS DIVERGENTES (ACRs e DVRs)

[000162] O esquema acima descrito para a seleção de um Resíduo Candidato conta com um critério de seleção que o Resíduo Candidato está presente em uma sequência de consenso de uma segunda proteína exibindo uma propriedade melhorada ou desejável, ao mesmo tempo sendo diferente do resíduo correspondente em uma pluralidade de sequências de consenso das primeiras proteínas que não exibem a propriedade melhorada ou desejável, e onde as sequências de consenso das primeiras proteínas compartilham o mesmo resíduo. Desse modo, usando o exemplo de especificidade de CO_2 em Rubisco, o Resíduo Candidato será um resíduo que está presente na sequência de consenso de algas vermelha, e que é diferente do resíduo comum nas sequências de consenso tanto de plantas floríferas como de cianobactérias.

[000163] Outras soluções evolutivas parciais podem também ser expressadas em resíduos encontrados no fundo geral de Resíduos Variantes. Estas foram denominadas Resíduos Candidatos Alternativos e Resíduos Candidatos Divergente. Estas soluções parciais encontradas

na natureza podem ser usadas para estender o processo de selecionar Mutantes Candidatos, para produzir um fundo geral expandido de resíduos alternativos ou adicionais com os quais para influenciar os Resíduos-alvos.

[000164] Por exemplo, onde o resíduo selecionado da segunda proteína não for o resíduo de maior parte encontrado na sequência de consenso da segunda proteína mas do contrário for encontrado na frequência mais baixa em uma pluralidade das segundas proteínas, embora ainda diferente dos resíduos presentes em uma pluralidade de sequências de consenso das primeiras proteínas, e onde as sequências de consenso das primeiras proteínas compartilham o mesmo resíduo, o resíduo é denominado um Resíduo Candidato Alternativo (ACR). O ACR representa uma alternativa ao resíduo de consenso a uma posição de Resíduo de Candidato. Como para um Resíduo Candidato, um ACR tem que satisfazer ainda os critérios de seleção relacionados a influenciar um Resíduo-alvo relevante. Um propósito para introduzir ACRs pode se prover resíduos para enxerto que são suspeitos de ser associados a uma maior melhoria da propriedade desejável que o resíduo de maior parte da sequência de consenso da segunda proteína.

[000165] Desse modo, usando o exemplo de especificidade de CO_2 em Rubisco, o Resíduo Candidato Alternativo será um resíduo que é expressado por pelo menos uma das algas vermelhas que contribuem para a sequência de consenso de algas vermelhas mas que não seja o resíduo de maior parte que está contido na sequência de consenso, e que seja diferente do resíduo comum nas sequências de consenso tanto de plantas verdes como de cianobactérias. As espécies algais vermelhas *Griffithsia monolis* têm uma taxa catalítica mais alta comparada com outras espécies algais vermelhas embora mantendo a especificidade alta típica das algas vermelhas (Whitney et al, 2001), e con-

sequentemente a sequência de *Griffithsia monolis* pode ser usada como uma fonte de ACRs. A variação da sequência de Rubisco de domínio N-terminal entre *G. monolis* e à da sequência de consenso de algas vermelha e uma espécie algal vermelha típica (referência), *Galdieri partita*, é ilustrada na Figura 14. Esta mostra ACRs nos resíduos 36, 51, 54, 88 e 104. Obserque que o resíduo 36 é também um DCR.

[000166] Um Resíduo Covariante (CvR) pode ser identificado na sequência de uma segunda proteína de uma espécie particular, como estando na redondeza de um Resíduo Candidato Alternativo (ACR) e mostrando variação complementar para o ACR. Esta variação na posição do CvR que não está presente na sequência de consenso para a segunda proteína pode ser suspeita de refletir alterações complementares nas propriedades estruturais e/ou eletrostáticas do ACR e CvR. Estas alterações complementares podem constituir uma solução evolutiva parcial. Identificação dos CvRs fornece um meio para identificar posições de resíduos auxiliares que podem ser mutadas na primeira proteína para melhor acomodar as alterações feitas de ACRs transplantados.

[000167] Ainda outra solução evolutiva parcial pode também ser expressada em resíduos encontrados no fundo geral de Resíduos Variantes onde os resíduos de pelo menos três ramificações filogenéticas são examinados e o resíduo selecionado é prognosticado para influenciar pelo menos um Resíduo-alvo e é encontrado na sequência de consenso da segunda proteína de uma ramificação embora sendo diferente dos resíduos presentes nas sequências de consenso de pelo menos duas outras ramificações, e onde as sequências de consenso das primeiras proteínas são também diferentes na mesma posição. Um tal resíduo é denominado um Resíduo Candidato Divergente (DCR). Como para um Resíduo Candidato, um DCR tem que satisfazer ainda os critérios de seleção relacionados a influenciar um Resí-

duo-alvo relevante. Um propósito para introduzir um DCR pode ser prover resíduos para enxerto que são suspeitos de estar associados à propriedade melhorada ou desejável, mas que podem também ser esperados produzir uma maior variação das propriedades expressas quando substituídos em diferentes primeiras proteínas (hospedeiro) da pelo menos duas ramificações filogenéticas que seriam esperadas se a substituição estivesse com um Resíduo Candidato.

[000168] Desse modo, usando o exemplo de especificidade de CO_2 em Rubisco, o Resíduo Candidato Divergente pode ser um resíduo que é expressado na sequência de consenso de algas vermelha, mas que é diferente do resíduo expresso nas sequências de consenso de plantas floríferas e na sequência de consenso de cianobactérias, enquanto ao mesmo tempo resíduos das sequências de consenso de plantas floríferas e de cianobactérias também diferem nesta posição. Seis Resíduos Candidatos Divergentes identificados como capazes de influenciar os Resíduos-alvos na Região 1 (ASN123, Glu60 e TYR20) são identificados através de sombreamento cinza no alinhamento na Figura 7, e são mostrados graficamente na Figura 13. Tabela 2 mostra a variação no resíduo entre plantas floríferas e cianobactérias nestas seis posições.

[000169] TABELA 2. Tabela de composição do resíduo (em porcentagem) para cianobactérias, plantas floríferas e algas vermelhas em 20 Resíduos Candidatos atuais (incluindo três do domínio C-terminal de LSU par) e 6 sítios de Resíduo Candidato Divergente. Número das sequências usadas é mostrado entre parênteses. Os DCRs estão ao fundo da tabela (36, 86, 116, 117, 138, 140).

Res No	**Cianobactérias (11)**		**Plantas floríferas (134)**		**Algas vermelhas (9)**		**Substituição em PCC6301 de *Synechococcus* através do resíduo em**			**Sub-região**
	Res	**%**	**Res**	**%**	**Res**	**%**	**algas vermelhas**	**plantas floríferas**	**cianobactérias**	
18	K Q	73 27	K	100	I -	67 33	K18Q		K18Q	1A
19	D E	81 18	D E	92 8	P -	67 33	D19E	D19E	D19E	1A

23	T	100	T N	99 1	G	100	T23G	T23N		1A
25	Y W	82 18	Y H	98 2	W	100	Y25W	Y25H		
51	E D	91 9	E	100	V I	56 44	D51V D51I		D51E	
54	G A	55 45	G R	99 1	S A	78 22	G54A G54S	G54R		1C
59	A	100	A	100	G	100	A59G			
64	G	100	G	100	A	100	G64A			
68	T	100	T A	99 1	V	100	T68V			1A
81	K	100	K	100	R	100	K81R			1A
84	C	100	C	100	A C	89 11	C84A			1C
87	I L V	82 99	I L	98 2	V	100	187 V	I87L		1C
88	E	100	E	100	D E	78 22	E88D			1C
104	P	100	P	100	D E	89 11	P104D P104E			1A
114	T	100	T	100	A	100	T114A			
118	T	100	T	100	A	100	T118A			
121	V	100	V	100	I	100	V121I			1B
271	T	100	T	100	V	100	T271V			
297	M	100	M	100	G	100	M297G			
300	V	100	V	100	T	100	V300T			
36	L I V	73 18 9	I	100	V I	89 11	L36V L36I			
86[+]	D R	63 18	H G D	57 20 16	K R	89 11	H86K H86R	H86G H86D	H86D H86R	1C
116	V M I	64 18 18	M L	99 1	L	100	I116L	I116M	I116V	1B
117	L	100	F L	99 1	T	100	L117T	L117F		1B
138	I L	82 18	L	100	M	100	I138M	I138L	I138L	1B
140	F I	82 18	I V S	97 2 1	L I	56 44	F140L F140I	F140V F140S		1B

[+] Para o resíduo 86, apenas aqueles resíduos alternados são exibidos com mais de 5% de ocorrência para cianobactérias e plantas floríferas.

[000170] Como mostrado na Tabela 2, a maioria dos CRs tem um resíduo alternativo em um ou mais dos 3 grupos (cianobactérias, plantas floríferas, algas vermelhas) embora na maioria dos casos o resíduo de consenso ocorra com >70% de frequência. Exceções são 59, 64, 81, 114, 118 e 121, e todos os três CRs (271, 297 e 300) nos segmentos de domínio C-terminais da LSU adjacente (Sub-região 1B). Porém,

a magnitude desta variação é de pouca significação como o conjunto de sequência para cada grupo não é filogeneticamente equilibrado. Do contrário esta variação deve apenas ser tomada como uma indicação aproximada de grau de conformidade com a definição de CR. O Mutante #4 mais forte (T23G/K81R) (vide Exemplos 7 e 11) não mostrou quase nenhuma variação. Similarmente, o mutante mais forte da subregião 1B, Mutante #8 (V 1211/M297G/V300T) (vide Exemplo 7), também não mostrou nenhuma variação dos 3CRs.

(iii) AGRUPAMENTO DE RESÍDUOS CANDIDATOS, RESÍDUOS CANDIDATOS ALTERNATIVOS E RESÍDUOS CANDIDATOS DIVERGENTES EM MUTANTES CANDIDATOS

[000171] Mais de um Resíduo Candidato (CR), Resíduo Candidato Alternativo (ACR) ou Resíduo Candidato Divergente (DCR), ou combinações dos mesmos, podem contribuir para alterações em uma região de interação contígua simples entre as sequências da primeira e segunda proteína. Por exemplo, um Resíduo de aminoácido não-variante simples pode interagir com dois CRs, dois ACRs, dois DCRs, ou uma combinação de dois resíduos derivados de dois destes grupos, de modo que ambas as interações alteradas podem afetar o mesmo Resíduo-alvo. Os dois CRs, dois ACRs, dois DCRs, ou combinação podem depois ser agrupados em um Mutante Candidato simples. Similarmente, um CR, ACR ou DCR dado pode contribuir para alterações em mais de uma região de interação contígua afetando um Resíduo-alvo. Desse modo, pode haver outros CRs e/ou ACRs e/ou DCRs com os quais um CR, ACR ou DCR dado possam ser agroupados para formar outros Mutantes Candidatos. Estes CRs, ACRs, DCRs agrupados ou combinações dos mesmos podem contribuir para um enzima de Mutante Candidato proposta onde cada CRs, ACRs, DCRs agrupados ou combinação dos mesmos é "enxertado" no hospedeiro, substituindo os resíduos correspondentes na sequência da primeira proteína (hospe-

deiro).

(iv) COMBINAÇÃO DE MUTANTES CANDIDATOS PARA PRODUZIR UM EFEITO CUMULATIVO

[000172] Se as mutações de resíduo de dois ou mais Mutantes Candidatos (grupos de umA ou mais CRs e/ou ACRs e/ou DCRs) afetarem a mesma unidade estrutural secundário abrigando um Resíduo-alvo e mutações de cada um destes Mutantes Candidatos são esperadas atuar em uma maneira coordenada alterando a interação, então os Mutantes Candidatos pode ser também combinados para formar um Mutante Candidato combinado simples. Por exemplo, se um Resíduo-alvo fizer parte de uma hélice com um Mutante Candidato interagindo com a extremidade N-terminal da hélice e outro Mutante Candidato interagindo com a extremidade C-terminal da hélice, com ambas as interações envolvendo adição de uma interação hidrofóbica forte indo da sequência da primeira proteína para a sequência da segunda proteína, então um Mutante Candidato combinado com ambos os Mutantes Candidatos enxertados na sequência da primeira proteína seriam prognosticados mostrar efeitos combinados em reposicionar a hélice.

(v) CLASSIFICAÇÃO DOS MUTANTES

[000173] Apesar da redução grande no número de resíduos para consideração para mutação que pode ser realizada pela seleção inicial dos Resíduos Variantes contra Resíduos-alvos para produzir a lista de Resíduos Candidatos (CRs), e opcionalmente, Resíduos Candidatos Alternativos (ACRs) e Candidatos Divergentes (DCRs), o número de Mutantes Candidatos prognosticados que poderiam resultar do agrupamento dos CRs, ACRs e DCRs ainda poderia ser grande. Por conseguinte, é útil classificar os Mutantes Candidatos com base em seu potencial prognosticado para mostrar intensificação da propriedade funcional desejada. Os Mutantes Candidatos classificados mais altos podem depois ser a primeira escolha para outras avaliações computa-

cionais e experimentais, desse modo minimizando esforço e custo. Como ilustrado no Fluxograma 2, vários princípios podem ser considerados na classificação;

(a) O(s) Resíduo(s)-Alvo(s) afetado(s) pelo Mutante Candidato. Resíduos-alvos diferentes podem ter níveis diferentes de participação na etapa de reação, e, consequentemente, importância relativa. Por exemplo, um Resíduo-alvo pode estar envolvido diretamente em uma transferência de prótons, enquanto outro pode ser apenas um doador/aceitante de ligação de hidrogênio passivo. Para um Resíduo-alvo com maior participação na etapa de reação, o efeito da mutação de um Mutante Candidato relacionado é prognosticado ser maior, e, consequentemente, o Mutante Candidato correspondente é classificado mais alto.

(b) Nível da interação dos CRs e/ou ACRs e/ou DCRs no Mutante Candidato. Se a alteração prognosticada nas interações associadas com mutações para um Mutante Candidato envolver formação, ou eliminação, de uma interação eletrostática ou hidrofóbica forte, então a classificação será mais alta.

(c) Número de alterações de interação acompanhando as mutações para um Mutante Candidato. Até mesmo se as mutações envolverem adição, ou remoção, de apenas interações fracas, as alterações totais podem envolver muitas tais interações dentro do ambiente de proteína de Resíduo-alvo, levando a um efeito cumulativo de rede grande na estrutura e eletrostáticas, e, consequentemente, produzem um efeito pronunciado na propriedade funcional. Consequentemente, tais mutações podem ser classificadas altamente.

(vi) PREDIÇÕES DE ENXERTO FILOGENÉTICO ESTENDIDO

[000174] Predições de enxerto filogenético estendido podem ser usadas para explorar a capacidade do procedimento de enxerto filoge-

nético para "aprender da experiência" através de interpretação dos resultados (sucessos e falhas) dos ciclos de aplicação do método de núcleo de predição e testagem dos Mutantes Candidatos para proteínas de interesse descritas acima. Embora o método de núcleo descrito acima focalizasse em reconhecer Resíduos Candidatos, Resíduos Candidatos Alternativos e Resíduos Candidatos Divergentes e agrupá-los em Mutantes Candidatos com base nas redes de interações destes resíduos, os resultados iniciais para Rubisco detalhados no Exemplo 3 mostraram que uma proporção destes Mutantes Candidatos eram ineficazes ou mesmo ligeiramente prejudiciais para a função da Rubisco, embora não grandemente. A estratégia de enxerto filogenético estendido permitiu este conhecimento acumulado, sucessos e falhas, serem reinterpretados e desenvolvido em um método também personalizado para a aplicação de proteína particular, e que pode continuar sendo refinado visto que resultados adicionais são obtidos. A interação entre a tecnologia de enxerto filogenético estendido e o método de núcleo descrito acima é mostrada nos Fluxogramas 1 a 3.

[000175] A estratégia de enxerto filogenético estendido tem três componentes principais. Primeiramente, refinamento do conceito de redes de interação dos Resíduos Candidatos, e/ou Resíduos Candidatos Alternativos e/ou Resíduos Candidatos Divergentes como a base do método de núcleo para agrupamento em Mutantes Candidatos. Exame dos sucessos e falhas iniciais dos mutantes de Rubisco da Região 1, e comparando *Synechococcus* em particular Mutante #6a e seus Mutantes de componente #4 e #1a (debatidos também nos Exemplos 3 e 4), sugeriu que uma estrutura melhorada para predição envolveria reconhecer que há "pontos de acesso" para adaptação evolutiva contendo as soluções evolutivas parciais para Rubisco melhorada, e que identificação destas "sub-regiãos mutáveis" fornece uma base melhor, ou adicional, para identificar Resíduos Candidatos, Resí-

duos Candidatos Alternativos e Resíduos Candidatos Divergentes que deveriam ser preferencialmente agrupados para formar Mutantes Candidatos para manipular as propriedades funcionais da Rubisco, ao invés de usar somente as redes de interação dentro das Regiões, como no método de núcleo. Desse modo, no método estendido, o foco de investigação foi mais em identificar regiões espaciais que em identificar as redes de interação dos resíduos.

[000176] Em segundo lugar, reinterpretação dos resultados iniciais para a mutantes de Rubisco da Região 1 dentro desta estrutura permitiu a identificação de volumes espacialmente contíguos da estrutura de proteína, chamadas Sub-regiões, contendo um subconjunto dos CRs da Região (incluindo ACRs e DCRs inicialmente identificados), e que poderiam ser prognosticados influenciar as propriedades de um Resíduo-alvo particular ou Resíduos-alvos preferencialmente ligados à Região. Identificação destes subconjuntos dos CRs, ACRs e DCRs da Região fornece um meio pelo qual eles podem ser preferencialmente agroupados para formar Mutantes Candidatos.

[000177] Em terceiro lugar, identificação das Sub-regiões com limites imprecisos e sobrepostos provê uma base para identificar CRs, ACRs, CvRs e DCRs adicionais que são prognosticados interagir com os subconjuntos de núcleo de CRs, ACRs e DCRs e que podem ser recrutados para os subconjuntos de núcleo para prover resíduos adicionais para agrupamento em Mutantes Candidatos.

[000178] Além disso, identificação das Sub-regiões como pontos de acesso de variação de sequência natural fornece o meio para identificar e explorar outros tipos de dados de diversidade de sequência, tais como Resíduos Variantes Espécie-específicos (SsVRs) que variam entre espécies estritamente relacionadas, e que podem representar soluções evolutivas parciais, tais como para adaptação a ambientes particulares tais como quente/seco ou frio/úmido. Usando Rubisco co-

mo um exemplo, a Figura 14 ilustra a variação de sequência natural entre duas espécies algais vermelhas e cinco espécies de plantas floríferas.

[000179] Por exemplo, três Sub-regiões da Região 1 da LSU de Rubisco foram identificadas como pontos de acesso mutáveis, e regiões rotuladas 1A, 1B e 1C, como mostradas nas Figuras 15 e 17. Como detalhado no Exemplo 4, os inventores usaram o método estendido para recrutar resíduos adicionais para conjunto de resíduos preferidos da Região 1 para mutação e formar Mutantes Candidatos adicionais prognosticados para serem melhorias dos melhores mutantes "guias" produzidos da aplicação do método de núcleo, detalhado no Exemplo 3.

[000180] Em resumo, o desenvolvimento do método de enxerto filogenético estendido fornece um meio para melhorar as propriedades funcionais dos mutantes "guias". O desenvolvimento do conceito de Sub-regiões como áreas de ponto de acesso e como uma estrutura para construir uma base de dados da sequência de proteína mapeada para dados experimentais de teste de predições, também posiciona a tecnologia de enxerto filogenética para melhor explorar os dados de diversidade de sequência. Por exemplo, a base de dados acima mencionada pode ser interrogada contra dados de diversidade de sequência para identificar resíduos que podem ser recrutados como resíduos variantes espécie-específicos (SsVRs) para possível inclusão nos Mutantes Candidatos. por exemplo, para Rubisco, os dados de sequência e, em alguns casos, os dados cinéticos estão disponíveis para organismos fotossintéticos que crescem sob ambientes variados ou atípicos, incluindo para espécies de planta de C3 relacionadas tais como aquelas encontradas nas Ilhas Baleáricas com tolerâncias diferentes à seca e temperatura (Galmes et al. 2005), o feijão de Marama africano sulista adaptado à seca (Parry et al, 2007), e algas vermelhas extre-

mofílicas *Cyanidiales* (Ciniglia et al., 2004). Alternativamente para Rubisco, em casos onde tal diversidade de sequência e outros dados funcionais estão disponíveis para plantas de plantação tais como trigo (Evans e Austin, 1986), a base de dados acima mencionada pode ser usada para identificar espécies naturais com propriedades funcionais melhoradas que podem ser usadas como germplasmo em procriação seletiva.

PRODUÇÃO DE PROTEÍNAS

[000181] Proteínas com pelo menos um Resíduo Candidato e/ou Resíduo Alternativo e/ou Resíduo Candidato Divergente, e opcionalmente incluindo outras substituições de CvRs e/ou SsVRs, da segunda proteína podem ser modeladas *in silico*, ou podem ser engenheiradas *in vitro* e/ou *in vivo*, por exemplo, através de mutagênese locodirigida de um polinucleotídeo que codifica a proteína e depois expressada em um sistema de expressão.

TRIAGEM DE PROTEÍNAS MUTANTE

[000182] Proteínas Mutantes Candidatas compreendendo a combinação de Resíduos-alvos com um ou mais Resíduos Candidatos e/ou Resíduos Candidatos Alternativos e/ou Resíduos Candidatos Divergentes, e opcionalmente incluindo outras substituições de Resíduos Covariantes (CvRs) e/ou Resíduos Variantes Espécie-específicos (SsVRs) são triadas para identificar depois aquelas proteínas Mutantes Candidatas que têm a dita propriedade funcional melhorada. O processo de triagem das proteínas Mutantes Candidatas pode usar qualquer uma ou mais de várias técnicas e pode compreender avaliação catalítica, avaliação bioquímica, e avaliação fisiológica.

EVOLUÇÃO DIRECIONADA

[000183] As proteínas da invenção podem ser modificadas através de evolução direcionada. Consequentemente, a função de uma proteína produzida pelos abordagens da invenção pode ser melhorada ou

do contrário modificada. Em geral, evolução direcionada pode envolver mutagenização de um ou mais modelos moleculares parentais e identificar qualquer molécula desejável entre as moléculas da progênie. Moléculas da progênie podem depois ser triadas para a propriedade desejada, avaliando, por exemplo, a atividade da molécula, a estabilidade da molécula, e/ou frequência de mutação na molécula. Moléculas da progênie com propriedades desejáveis podem ser depois selecionadas e outros ciclos de mutagênese e triagem executados. Abordagens pelos quais evolução direcionada pode ser executada são bem conhecidos na técnica. Abordagens exemplares incluem, entre outros, abordagens de evolução direcionada racional descrios no pedido de patente U.S. N°. 10/022.249; e Pedido de patente U.S. Publicado N°. US-2004-0132977-A1. Para uma descrição geral da metodologia experimental e técnicas envolvidas na evolução direcionada, referência pode ser feita a Sambrook et al., Molecular Cloning, A Laboratorial Manual 2ª ed., Cold Spring Harbor Laboratory Press, 1989.

[000184] Em uma modalidade da invenção, proteínas mutantes produzidas pelos abordagens da invenção podem ser usadas como um "ponto de partida" para evolução direcionada. Tomando o exemplo de uma proteína de Rubisco, proteínas Mutantes Candidatas proteínas observadas ter atividade enzimática melhorada podem ser selecionadas e também otimizadas através de evolução direcionada. Esta otimização pode facilitar, por exemplo, o alívio dos conflitos estéricos por recrutamento de SvRs e outros resíduos de variante de ocorrência natural que podem ser identificados como complementares para mutações nas proteínas Mutantes Candidatas. Proteínas Mutantes Candidatas com atividade enzimática melhorada podem desse modo servir como um ponto de partida novo para evolução direcionada como elas têm potencial diferente para explorar espaço da sequência comparado com a proteína de Rubisco do tipo selvagem. Sistemas adequados pa-

ra evolução direcionada de Rubisco incluem, mas não são limitados àqueles empregando o MMI de cepa de *E. coli* como recentemente foi relatado (Mueller-Cajar et al., 2007).

PROTEÍNAS DE RUBISCO

[000185] As proteínas geradas de acordo com a invenção incluem equivalentes funcionais, variantes, fragmentos ativos e proteínas de fusão. Para evitar dúvidas, o seguinte é incluído dentro do escopo da invenção: equivalentes funcionais dos fragmentos ativos e proteínas de fusão; fragmentos ativos dos equivalentes funcionais e proteínas de fusão; e proteínas de fusão compreendendo um equivalente funcional ou fragmento ativo.

[000186] O termo "fragmento" refere-se a uma sequência de ácido nucleico ou de polipeptídeo codifica um constituinte ou é um constituinte de proteína de comprimento total. Em termos do polipeptídeo, o fragmento possui atividade biológica qualitativa em comum com a proteína de comprimento total. Um fragmento biologicamente ativo de uso de acordo com a presente invenção pode tipicamente possuir pelo menos cerca de 50% da atividade da proteína de comprimento total correspondente, mais tipicamente pelo menos cerca de 60% de tal atividade, mais tipicamente pelo menos cerca de 70% de tal atividade, mais tipicamente pelo menos cerca de 80% de tal atividade, mais tipicamente pelo menos cerca de 90% de tal atividade, e mais tipicamente pelo menos cerca de 95% de tal atividade.

[000187] Abordagens de medir a identidade da sequência de proteína são bem conhecidos na técnica e serão entendidos por aqueles de habilidade na técnica que no contexto presente, identidade de sequência é calculada em base da identidade de aminoácido (às vezes referida como "homologia rígida"). Identidade de sequência é calculada após alinhar as sequências. Os inventores usaram o programa de ClustalW (Thompson et al, 1994) fornecido dentro do Editor de Alinha-

mento de Sequência BioEdit (Runner, 1999) para alinhar as sequências. Há vários pacotes de software livre-para-usar e particulares disponíveis que executam alinhamentos de sequência e rendem eficazmente os mesmos resultados. A identificação de Resíduos Variantes pode também ser executada colhendo sequências de Rubisco, usando, por exemplo, pesquisas de BLAST de um ou mais grupos filogenéticos que diferem na propriedade cinética selecionada para melhoria, e alinhando-as junto à dita primeira sequência, usando, por exemplo, CLUSTALW.

[000188] Os equivalentes funcionais, fragmentos ativos e proteínas de fusão da invenção retêm a habilidade da proteína (SEQ ID NO: 23 para PCC7942 de *Synechococcus* sp. e SEQ ID NO: 72 para tabaco) para atuarem como uma enzima de Rubisco com eficiência melhorada. Porém, pessoas versadas na técnica poderão inventar ensaios ou meios para avaliar a atividade enzimática.

[000189] Proteínas funcionalmente equivalentes são pretendidas de acordo com a invenção, portanto, incluem mutantes (tais como mutantes contendo substituições, inserções ou deleções de aminoácidos). Tais mutantes podem incluir proteínas em que um ou mais dos resíduos de aminoácido são substituídos com um resíduo de aminoácido conservador ou não-conservador e tal(is) resíduo(s) de aminoácido substituído(s) podem ou não ser codificados pelo código genético.

[000190] Particularmente preferidas são proteínas em que vários, isto é, 30 e 50, entre 20 e 30, entre 15 e 20, entre 10 e 15, entre 5 e 10, 1 e 5, 1 e 3, 1 e 2 ou apenas 1 aminoácido é/são substituído(s), deletado(s) ou adicionado(s) em qualquer combinação. Proteínas "mutantes" também incluem proteínas em que um ou mais dos resíduos de aminoácido incluem um grupo de substituintes.

[000191] Tais fragmentos podem ser "soltos", isto é, não partem, ou não são fundidos, de outros aminoácidos ou proteínas, ou eles podem

ser compreendidos dentro de uma proteína maior da qual elas formam uma parte ou região. Quando compreendido dentro de uma proteína maior, o fragmento da invenção em uma modalidade forma uma região contínua simples. Adicionalmente, vários fragmentos podem ser compreendidos dentro de uma proteína maior simples.

[000192] Em uma modalidade da invenção é fornecida uma proteína de fusão compreendendo uma proteína da invenção fundida em um peptídeo ou outra proteína, tal como uma marcação que pode ser por exemplo, bioativa, radioativa, enzimática ou fluorescente, ou um anticorpo.

[000193] Por exemplo, é frequentemente vantajoso incluir uma ou mais sequências de aminoácido adicionais que podem conter sequências secretórias ou líderes, pró-sequências, sequências que auxiliam na purificação, ou sequências que conferem estabilidade de proteína mais alta, por exemplo, durante a produção recombinante. Alternativa ou adicionalmente, a proteína madura pode ser fundida com outro composto, tal como um composto para aumentar a meia-vida da proteína (por exemplo, polietileno glicol).

FUNÇÕES DA ENZIMA

[000194] Em outra modalidade a proteína gerada de acordo com a invenção é uma enzima. Onde a proteína for uma enzima, a função pode ser a catálise de pelo menos uma reação química. Em outras modalidades, a função pode ser estrutural (por exemplo, servindo como uma proteína citoesquelética). A função pode envolver o transporte ativo ou passivo de uma substância dentro da célula ou entre o interior e exterior da célula, ou entre compartimentos diferentes dentro da célula, ou entre regiões diferentes do organismo, por exemplo, onde a proteína estiver envolvida em um canal ou um poro da membrana, ou a proteína estiver envolvida no tráfego de materiais para compartimentos celulares específicos ou a proteína atuar como uma chaperona ou

um transportador. A função pode estar envolvida com interações de ligando/receptor, por exemplo, onde a proteína for um fator de crescimento, uma citocina, um neurotransmissor ou um ligante intracelular ou extracelular, ou a proteína for um receptor para o fator de crescimento, citocina, neurotransmissor ou o ligante intracelular ou extracelular.

[000195] Onde a proteína for uma enzima, a enzima pode estar envolvida no catabolismo ou metabolismo. A enzima pode estar envolvida na síntese de pelo menos um produto. A enzima pode estar envolvida no desarranjo de pelo menos um substrato. A enzima pode estar envolvida na modificação química de pelo menos um substrato, por exemplo, a adição ou deleção de um ou mais grupos de fosfato de uma molécula.

[000196] As enzimas podem ser adequadas para o uso, por exemplo, na degradação de pesticidas, e resíduos detergentes, para extração mineral, ou para processos químicos "volumosos" ou fine finos, tais como amilases. As enzimas podem também ser adequadas para o uso em aplicações médicas, e em particular podem ser usadas por minimizar as alterações na estabilidade biológica e físico-química.

[000197] As enzimas podem ter propriedades especificamente engenheiradas, por exemplo, a habilidade para executar otimamente em uma faixa de temperatura desejada, uma especificidade de substrato mais estreita, mais ampla ou alterada, ou a habilidade para impedir a produção e/ou liberação de tóxico ou subprodutos potencialmente tóxicos. A enzima pode ser reprojetada de modo que ela seja um catalisador eficiente para uma reação secundária da enzima do tipo selvagem usando seu substrato natural ou um substrato alternativo para produzir um produto diferente.

PURIFICAÇÃO DAS PROTEÍNAS DE RUBISCO

[000198] A invenção fornece um método de purificar uma proteína de

Rubisco produzido de acordo com os abordagens da invenção. A holoenzima da forma funcional da Rubisco de organismos eucarióticos (Rubisco de forma I) é um hexadecâmero feito de 8 subunidades grandes (LSUs) e 8 subunidades pequenas (SSUs), e requer enovelamento e montagem da enzima corretamente para chaperonas apropriadas. *E. coli* é o hospedeiro microbiano amplamente usado para expressar o DNA recombinante e as proteínas. Quando o óperon que codifica para os genes de Rubisco (*rbc*LS e *rbc*SS) de PCC7942 de *Synechococcus* sp. for expressado em *E. coli* ambas as subunidades são abundantemente sintetizados, porém apenas cerca de 1 a 5% da LSUs expressos são corretamente enovelados e montados em forma funcional com a quantidade de Rubisco funcional acumulando a ~1 a 3% (p/p) da proteína solúvel de *E. coli*.

[000199] Para superar estas dificuldades, um sistema recentemente adaptado (Baker et al., 2005, os conteúdos inteiros deste são aqui incorporados por referência) pode ser usado para purificar proteínas de Rubisco nativas ou mutantes. Neste caso, o sistema acima mencionado foi usado para a purificação de Rubisco de PCC7942 de *Synechococcus* sp. expresso em *E. coli*. A primeira etapa do método de purificação envolve fundir em um primeiro vetor a sequência de codificação com uma Sequência de ubiquitina (Ub) marcada com H_6 (H_6Ub) à extremidade 5' de um gene *rbc*SS. Um hospedeiro é depois cotransformado com o primeiro vetor e segundo vetor que codifica para a subunidade grande nativa (ou mutada) e subunidade pequena da proteína de Rubisco, e expressão da proteína de Rubisco e vetores são depois induzidos, produzindo todos os três peptídeos de subunidade de Rubisco (isto é, LSU, SSU e H_6UbSSU). Alguns são montados nos hexadecâmeros de Rubisco funcionais compostos de 8xLSU núcleos octaméricos e razões diferentes de SSU (no máximo 8) e H_6UbSSU. A proteína de Rubisco é purificada com base na expressão do marcador de

H_6 fundida na subunidade pequena de Rubisco. Esta purificação pode ser executada, por exemplo, usando técnicas de cromatografia tais como cromatografia de afinidade de metal. Os fragmentos de Ub podem ser depois removidos da Rubisco usando, por exemplo, uma protease Ub-específica.

[000200] A presente invenção agora será descrita com referência aos exemplos específicos que não deveriam ser interpretados como limitando o escopo da invenção de qualquer forma.

EXEMPLOS

EXEMPLO 1

IDENTIFICAÇÃO DOS RESÍDUOS-ALVOS - QUÍMICA COMPUTACIONAL

[000201] Um estudo computacional do mecanismo completo de carboxilação de Rubisco e mecanismo de oxigenação complementar usando cálculos de QM *ab initio* em um complexo de fragmento do sítio ativo estendido provê a base para a estratégia.

[000202] Este modelo de complexo de fragmento compreende fragmentos da maioria dos resíduos de aminoácido de sítio ativo que ou foram estabelecido ou debatidos ter um papel fundamental na série de reações que são catalisadas no sítio ativo da Rubisco. Ele contém todos os resíduos diretamente coordenados para Mg_2^+ ou interagindo com o centro reativo do substrato. Este modelo de complexo de fragmento foi construído das coordenadas da estrutura de cristal com 8ruc de código PDB (estrutura de cristal do complexo de Rubisco ativado com Mg_2^+ e 1,5-bisfosfato de 2-carboxiarabinitol (2CABP)). Inicialmente, o modelo de complexo do fragmento da Rubisco com 1,5-bisfosfato de 2-carbóxi-3-cetorarabinitol (2C3KABP), um análogo estrutural de 2CABP e o intermediário da reação atual produzido durante a atividade de carboxilase da Rubisco, foi construído das coordenadas cristalográficas. Como mostrado na Figura 2, o modelo de fragmento de 77

átomos, nomeado FM20, compreende espécies de fragmento molecular para representar os resíduos de aminoácido LYS175, LYS177, ASP203, GLU204, KCX201 (LYS201 carbamilado), HIS294, e LYS334 e o fragmento de 4 carbonos da forma de enodiolato do substrato RuBP, mais as moléculas de água e de dióxido de carbono. Esta estrutura foi otimizada usando o pacote de química quântica Gaussian 03 (Frisch et al., 2004). Geometrias supostas para todas as outras espécies de reação na via de reação de carboxilase (mostrada nas Figuras 1 e 3) foram geradas modificando esta geometria otimizada, e suas estruturas de energia ótimas foram depois obtidas, também usando o pacote Gaussian 03. Em geral, várias possíveis estruturas para cada espécie diferindo, por exemplo, por padrão de ligação de H ou orientação dos átomos, foram consideradas. Por conseguinte, os papéis destes grupos foram examinados na adição de gás e etapas subsequentes das reações de carboxilase e de oxigenase levando a predições mais confiantes, em particular, para definição dos grupos envolvidos diretamente na adição de gás (CO_2 ou O_2) etapas e a energética relativa das duas etapas.

[000203] A análise é focalizada completamente na subunidade grande (LSU). Embora a maioria das Rubiscos, incluindo plantas verdes, algas e cianobactérias, sejam proteínas multiméricas complexas (hexadecaméricas) que consistem em 8 subunidades grandes (LSU; ~475 resíduos) e 8 subunidades pequenas (SSU; ~140 resíduos), a química do sítio ativo é conduzida por uma região de proteína que consiste em um dímero de LSUs apenas, com 8 tais sítios ativos de dímero na proteína hexadecamérica. Uma metade do domínio C-terminal (Tim-barrel) de uma LSU contém a maioria dos resíduos de sítio ativo, enquanto uma região menor do domínio N-terminal da LSU adjacente completa o sítio ativo. Porém, predições que surgem mais em geral de estudos da bioinformática sugerem que outras regiões possam estar

envolvidas na modulação da química, por exemplo, contatos de inter-subunidades (LSU-LSU ou LSU-SSU).

[000204] Figura 1 mostra um esquemático do mecanismo de reação de carboxilase proposto com base nos cálculos de QM. As setas curvadas indicam o fluxo de elétrons, significando a formação de ligação e eventos de quebra de ligação que levas às espécies de reação sucessivas. O esquemático define a participação dos resíduos de aminoácido em cada etapa de reação, e a etapa de adição de gás (carboxilação), em particular. Os resíduos prognosticados terem um papel na etapa de adição de gás compreendem o conjunto de Resíduos-alvos previamente definido. Figuras 3 e 4 fornecem geometrias e energias relativas, respectivamente, das espécies de reação da carboxilase.

[000205] As características mais significativas do mecanismo de reação proposto são debatidas daqui por diante.

[000206] Primeiramente, os inventores fizeram a descoberta surpreendente que H_2O[Mg] não é deslocado da coordenação de Mg por CO_2 durante carboxilação. A molécula de água de fato ajuda a ligar CO_2 ao sítio ativo e contribui para a estabilidade do produto carboxilado e o TS correspondente. A mesma molécula de água atua como a água de hidratação na etapa posterior. Previamente este papel de hidratação tinha sido atribuído a uma molécula de água encontrada na redondeza da esfera de coordenação.

[000207] Os inventores fizeram a descoberta surpreendente também que o átomo de O_2 permanece não-protonado no intermediário de enodiolato, apesar das expectativas dos princípios químicos gerais que necessitariam ser desprotonados para direcionar carboxilação exclusivamente para C2, do que para C3. Cargas atômicas derivadas de ESP também mostram que O3 é mais negativo que O2. Porém, este resultado inesperado é explicado pela observação das ligações de hidrogênio fortes entre LYS 175 e KCX201 protonado com O2, que efi-

cazmente impede O2 de direcionar carboxilação para C3. Adicionalmente, como LYS334 é H-ligado ao grupo fosfato de P1 na enzima, sua interação com o substrato CO_2 seria rompida se a carboxilação de C3 acontecesse, não deixando nenhum escopo para estabilização do TS e intermediário correspondente. Características também do mecanismo de reação elucidadas pelos inventores são como segue. KCX201 tem um papel direto apenas na reação de enolização inicial e permanece em um estado protonado após enolização. KCX201 e LYS175 têm um papel em impedir carboxilação de C3 neutralizando parcialmente a carga negativa em O2. HIS294 tem um papel significativo na catálise de multi-etapas de Rubisco. Ele transporta o próton entre Nε e O_3, modulando o comprimento de ligação de C3-O3 apropriadamente. GLU204 ativa a molécula de água coordenada por Mg para hidratação separando seu próton. Desse modo, carboxilação e hidratação ocorrem na mesma face do intermediário de enodiolato. O próton de H3 é eventualmente transferido para O2 apenas após a formação do intermediário de ácido *aci* (VII). A carga no intermediário de ácido *aci* é estabilizada por LYS175 e LYS334. LYS175 assegura protonação estereospecífica do carbono C2 para render os produtos finais. LYS334 compartilha seu próton com LYS175.

[000208] Com base nas descobertas acima, os inventores identificaram dois resíduos de aminoácido, um atuando como uma base o outro como um ácido, para cada etapa de reação. Para a etapa de adição de gás, HIS294 atua como uma base separando o próton O3, enquanto LYS334 é um ácido doando um próton para estabilizar o grupo carboxilato formado pela adição de CO_2. Alteração no ambiente estérico ou eletrônico destes dois resíduos fundamentais, ou qualquer outro resíduo que é structural ou quimicamente (através de interações eletrostáticas) ligado a eles, causaria impacto na especificidade da enzima e provavelmente também afetaria k_{cat}.

[000209] Resíduos TYR20 GLU60, ASN123 foram também identificados como sendo cruciais para orientação apropriada das moléculas de gás com relação ao substrato antes da adição, e para a estabilidade do aduto de gás e as estruturas de estado de transição correspondentes. Estes cinco resíduos compreendem um grupo inicial de Resíduos-alvos para mais examinação.

EXEMPLO 2

ANÁLISE FILOGENÉTICA PARA IDENTIFICAR RESÍDUOS VARIANTES CONTENDO RESÍDUOS DETERMINANTES DE ESPECIFICIDADE

[000210] Sequências de LSU de Rubisco de filos disponíveis dos organismos fotossintéticos foram colhidos para análise filogenética de bases de dados publicamente disponíveis em NCBI (www.ncbi.nlm.nih.gov/) e JGI (http://img.jgi.doe.gov/cgi-bin/pub/main.cgi) executando pesquisas BLAST de proteína (Altschul et al., 1997) usando a sequência de LSU de Rubisco de espinafre como a sequência de consulta. Como as sequências de LSU são tão distintivas, e a conservação relativamente alta comparada com a maioria das classes mais homólogas de proteínas em tais distâncias evolutivas amplas, elas são fáceis de identificar e outros programa de pesquisa livre-para-usar ou privado funcionaria igualmente bem.

[000211] Alinhamento das sequências de LSU de Rubisco extraídas de organismos fotossintéticos pertencendo a treze filos diferentes abrangendo algas vermelhas, cianobactérias, glaucofitos e plantas (10 filos) foi realizado usando software de ClustalW (Thompson et al, 1994) dentro do Editor de Alinhamento de Sequência BioEdit (Hall, 1999) para avaliar sua diversidade. Um alinhamento é mostrado na Figura 5; em casos onde hão mais de uma sequência de um filo simples, uma sequência de consenso de 50% foi usada para eficiência. As sequências de consenso foram obtidas usando o servidor em

http://coot.embl.de/Alignment//consensus.html. Qualquer outra ferramenta de alinhamento particular ou livre-para-usar poderia ser usada para o mesmo propósito devido ao fato que a conservação geral destes Rubiscos é tão alta que os mesmos resultados de alinhamento seriam eficazmente obtidos.

[000212] Figura 5 mostra quão diversas as sequências de LSU de Rubisco de filogenias distantes podem ser. Desse modo, embora informação sobre soluções evolutivas parciais para adaptação de Rubiscos a ambientes específicos seja embutida nesta diversidade, é impossível analisar esta informação diretamente para identificar resíduos responsáveis por adaptações específicas, tais como especificidade aumentada.

[000213] Para solucionar este problema, os inventores desenvolveram o método de enxerto filogenético baseado em hipótese e o aplicaram em identificação dos resíduos de Rubisco que já representam soluções evolutivas parciais naturais para especificidade intensificada ($S_{c/o}$) - é bem conhecido que organismos fotossintéticos diversos exibem caracteristicamente diferentes valores para $S_{c/o}$. A maior parte das plantas terrestres possui um valor de $S_{c/o}$ típico de 80, enquanto as algas vermelhas são conhecidas ter o fator de especificidade mais alto (~160). O fator de especificidade para cianobactérias tem um valor modesto de cerca de 40.

[000214] Como cianobactérias são um antepassado comum tanto das plantas terrestres como das algas vermelhas, e visto que as plantas terrestres divergiram das cianobactérias mais cedo em evolução comparadas com as algas vermelhas (http://www.geocities.com/we_evolve/Plants/ cloroplast.html) a solução evolutiva parcial para especificidade intensificada embutida nas alterações do resíduo de aminoácido em algas vermelhas pode ser revelada parcialmente por comparação das Sequências de LSU de Rubisco

destes três grupos. Figura 6 mostra o alinhamento das sequências de consenso de 50% de LSUs de Rubisco de algas vermelhas, cianobactérias e plantas floríferas (verdes) previamente mostradas na Figura 5, mas agora mostradas com sombreamento cinza para destacar os 134 resíduos nas algas vermelhas que diferem daqueles em comum em cianobactérias e plantas floríferas, isto é, os Resíduos Variantes. Como há um acordo geral que alterações de apenas um resíduo não resultarão na melhoria substancial em especificidade de Rubisco (do contrário a evolução ter sido capaz de explorar o espaço da sequência razoavelmente de modo fácil para otimizar a especificidade em organismos dados), o número de mutações múltiplas, por exemplo, de 2-10 resíduos, que necessitaria ser testado para explorar o espaço da sequência reduzido até mesmo destes 134 resíduos para identificar os resíduos determinantes de especificidade é impossivelmente grande. Fazendo uso desta lista de Resíduos Variantes, variações que não são relevantes à propriedade da enzima de interesse são filtradas (isto é, especificidade).

EXEMPLO 3

IDENTIFICAÇÃO, AGRUPAMENTO E CLASSIFICAÇÃO DE RESÍDUOS CANDIDATOS, E PREDIÇÃO DE MUTANTES CANDIDATOS USANDO MÉTODO DE ENXERTO FILOGENÉTICO DE NÚCLEO

[000215] Para identificar os resíduos determinantes de especificidade que respondem pela especificidade aumentada de Rubiscos algais vermelhas, os inventores usaram os entendimentos mecanísticos enzimáticos dos cálculos de QM descritos no Exemplo 1 para desenvolver um procedimento para selecionar um subconjunto dos 134 Resíduos Variantes identificados no Exemplo 2. Os Resíduos Variantes foram selecionados, identificados no Exemplo 2, junto aos Resíduos-alvos, identificados no Exemplo 1 para ter um papel funcional na etapa de adição de gás para identificar Resíduos Candidatos. Vários dos

Resíduos Candidatos selecionados foram também Resíduos Candidatos Alternativos. Além disso, vários Resíduos Candidatos Divergentes foram também selecionados. Este usou um procedimento com base no princípio geral que alterações evolutivas de resíduos que modificaram as propriedades destes Resíduos-alvos alterariam sua especificidade e a eficiência cinética. Desse modo, qualquer Resíduo Variante que tem o potencial para eletronicamente afetar os Resíduos-alvos ou estruturalmente, e, consequentemente, causar uma alteração na propriedade funcional da enzima, neste caso especificidade e eficiência cinética, é julgado fazer parte da solução evolutiva parcial para optimização da propriedade.

[000216] O procedimento de seleção foi aplicado para identificar tais Resíduos Candidatos, Resíduos Candidatos Alternativos e Resíduos Candidatos Divergente. Seleção foi realizada por análise visual e comparação das diferenças em interações de interresíduos com Accelrys Discovery Studio vl.5.1 (Accelrys Software Inc., San Diego, CA, 2005), usando as estruturas de cristal de Rubiscos de espinafre (código de PDB: 8ruc) e *Galdieria partita* (código de PDB: IBWV). A estrutura de cristal de Rubisco de espinafre foi usada em vez de Rubisco de PCC6301 de *Synechococcus* sp. (código de PDB: 1RB1) porque a resolução das estruturas disponíveis é superior para a estrutura de Rubisco de espinafre. Porém, os inventores responderam por alterações de resíduo entre LSUs de espinafre e Rubiscos de PCC6301 de *Synechococcus* sp. ao analisar as interações de interresíduos, e usaram a estrutura de *Synechococcus* também para referência.

[000217] Três dos cinco Resíduos-alvos, TYR20, GLU60 e ASN123 estão no domínio N-terminal da LSU de Rubisco, enquanto HIS294 e LYS334 estão no domínio C-terminal que forma a estrutura de Timbarrel (Figura 8). Para o propósito de classificar os grupos de CRs, ACRs e DCRs que podem formar subconjuntos de resíduos das solu-

ções evolutivas parciais, os Resíduos-alvos foram classificados como pertencendo a três Regiões diferentes. Resíduos TYR20, GLU60 e ASN123 interagem coletivamente com o grupo carboxilato do intermediário formado após adição de CO_2 (Figuras 8, 9 e 15). As unidades estruturais secundárias contendo estes três resíduos também entrelaçam com outras hélices e folhas do domínio N-terminal (Figuras 8, 9 e 15). Como evidente na Figura 15, ASN123, GLU60 e TYR20 estão presos a três regiões separadas de estrutura secundária do domínio N-terminal e suas cadeias laterais estendem-se para o sítio ativo em uma constelação tripartida. Consequentemente, estes três Resíduos-alvos são classificados como resíduos da Região 1. HIS294 está na fita β4, que é uma parte do TIM barrel; a região HIS294 circunvizinha é definida como Região 2. Embora LYS334 esteja na redondeza dos resíduos de HIS294 e da Região 1, ele está na 'alça' 6 da LSU (Figura 8) e na forma aberta da LSU ele está separado dos outros Resíduos-alvos; consequentemente, a região contendo LYS334 é classificada como Região 3.

[000218] A análise inicial e projeto de Mutante de Candidato são ilustrados por consideração dos Resíduos Variantes na Região 1. Como o domínio N-terminal da LSU de Rubisco é um domínio compacto (Figuras 9 e 15), considerações estruturais de proteína sugerem que variações do resíduo de aminoácido na maioria do domínio possam influenciar TYR20, GLU60 e ASN123 para os Resíduos-alvos. O procedimento de seleção foi iniciado mapeando todas as interações eletrostáticas e hidrofóbicas entre cada um dos Resíduos Variantes e todos os outros resíduos na LSU em ambas as estruturas de cristal. Resíduos variantes tendo interações equivalentes em ambas as estruturas de cristal Foram ignorados. As interações equivalentes compreendem interações que envolvem um átomo de cadeia principal de um Resíduo Variante principalmente, e tais interações ocorrem em grande parte den-

tro de uma hélice ou entre dois β-fitas. As interações foram também consideradas equivalentes se os Resíduos Variantes em espinafre (ou PCC6301 de *Synechococcus* sp.) e *Galdieria partita* diferirem apenas no comprimento da cadeia lateral e os átomos de interação são do mesmo tipo químico e dentro de aproximadamente a mesma distância em ambas as estruturas de cristal. Similarmente, as interações que envolvem valina, leucina e isoleucina foram consideradas equivalentes se apenas um grupo metila da cadeia lateral estiver envolvido na interação e a distância da interação não diferir significativamente entre as duas estruturas de cristal. Este processo removeu alguns dos Resíduos Variantes de maior consideração como Resíduos Candidatos potenciais (ou Resíduos Candidatos Divergente). O restante das interações que envolvem Resíduos Variantes são depois analisados para seu potencial afetar os Resíduos-alvos.

[000219] Um Resíduo Variante cuja cadeia lateral interage com a cadeia principal ou cadeias laterais dos Resíduos-alvos ou com resíduos que fazem parte das unidades estruturais secundários abrigando os Resíduos-alvos foi considerado um Resíduo Candidato (ou Resíduo Candidato Divergente). Como um exemplo, Resíduo-alvo GLU60 está na extremidade C-terminal da hélice αB e o Resíduo Variante ILE51, na extremidade N-terminal desta hélice, interage com outro Resíduo Variante TRP25 na LSU de *Galdieria partita* (vide Figura 14 e Tabela 2), mas esta interação está ausente nas LSUs de PCC6301 de espinafre e de *Synechococcus* sp. "Enxertando" estes Resíduos Variantes de *Galdieria partita* em PCC6301 de *Synechococcus* sp. é prognosticado ter um efeito na orientação da hélice αB e, consequentemente, em GLU60. Ambos estes Resíduos Variantes 51 e 25 foram, desse modo, considerados como Resíduos Candidatos. As posições destes resíduos estão mostradas na estrutura de LSU na Figura 9. Neste caso, é útil agrupar estes dois Resíduos Candidatos que compartilham uma

região de interação contígua em uma mutação simples, como é provável de eles formarem uma parte coordenada da solução evolutiva parcial. Este agrupamento mais simples é numerado Mutante Candidato #1a na Tabela 3. Tais Resíduos Candidatos Mutantes representam mutações para substituição dos resíduos (enxerto) da sequência de *Galdieria partita* sobre a sequência de PCC6301 de *Synechococcus* sp., com a vista para transferir características de especificidade mais alta em Rubisco de PCC6301 de *Synechococcus* sp.

[000220] Antes do enxerto, os grupos de Resíduos Candidatos podem ser também combinados em grupos que podem atuar também coordenando e amplificando o efeito perturbativo em um Resíduo-alvo dado ou Resíduos-alvos. Tal agrupamento estendido é útil se dois grupos diferentes de Resíduos Candidatos afetarem a mesma unidade estrutural secundária. Por exemplo, ambos os grupos do Resíduo Candidato (25, 51) e (54, 84, 87) afetam a hélice αB que abriga GLU60, através de duas interações diferentes. Consequentemente, um dos mutantes prognosticados compreendeu estes grupos combinados (mostrado como Mutante #7a na Tabela 3).

[000221] Os Mutantes Candidatos potenciais enxertados são também avaliados para verificar por novas interações estéricas desfavoráveis introduzidas mediante enxerto dos resíduos de *Galdieria partita* em PCC6301 de *Synechococcus* sp. Tais interações estéricas indesejáveis podem ser retificadas adicionando mutações complementares espaciais ao Mutante Candidato, ou poderiam ser investigados através de simulações de MD para avaliar se relaxamento estrutural para aliviar tais contatos ruins é energicamente acessível.

[000222] A etapa final no procedimento de predição de Mutante Candidato é classificar os mutantes potenciais enxertados para desenvolver uma lista classificada para o uso na priorização de testagem experimental ou pré-triagem *in silico* computacional detalhada. A clas-

sificação reflete o grau esperado ao qual as mutações combinadas nos Mutantes Candidatos individuais são esperadas alterar a propriedade funcional, neste exemplo na direção para melhoria de especificidade influênciando a química, e química relativa, das etapas de adição de gás para CO_2 e O_2. Classificação depende de vários parâmetros tais como o Resíduo-alvo afetado pelo grupo de resíduo Candidato, a resistência das interações alteradas do grupo de resíduo Candidato dentro da Região de resíduo-alvo, e o número de tais interações para cada grupo de resíduo Candidato.

[000223] Além da sequência de Rubisco de *Galdieria partita*, a sequência de outra espécie algal vermelha *Griffithsia monolis*, que é conhecida ter um k_c^{cat} melhor que *Galdieria partita*, foi considerado na análise. Para o conjunto inicial de Resíduos Candidatos selecionados, dois resíduos (51 e 54) mostram diferenças entre *G. partita* e *G. monolis*, isto é, eles são Resíduos Candidatos Alternativos. Mutantes Candidatos com ambas as variantes de resíduo nas posições 51 e 54 foram formados, como mostrados na Tabela 3 para #1, #5, #6, #7 e #13 de CM.

[000224] Como um exemplo, o conjunto de dezesseis grupos de resíduo Candidato mostrado na Tabela 3 foi prognosticado de análise inicial dos Resíduos-alvos da Região 1, formando 21 Mutantes Candidatos potenciais (com variantes *G. partita* e *G. monolis*). Para cada Mutante Candidato, a última coluna na Tabela 3 detalha a alteração estrutural prognosticada associada às mutações. Para alguns Mutantes Candidatos, estas alterações são explicadas graficamente em uma figura; a segunda coluna da tabela dá o número de figura para estes CMs.

[000225] Tabela 3 também mostra as classificações das prioridades para teste experimental. Tabela 3 mostra que dois Resíduos Candidatos Divergentes (36, 116 e 140; vide Tabela 2) foram selecionados

nesta análise inicial pelas razões resumidas para os Mutantes Candidatos relevantes (#9 e #10). Duas destas DCRs (36 e 140) também mostram diferenças entre *G. partita* e *G. monolis* (vide Tabela 2); para Mutantes Candidatos #9 e #10, a variante de Gm de 36 e a variante de Gp de 140 foram julgadas ser as mais promissoras para transplante em *Synechococcus*.

[000226] TABELA 3. Mutantes Candidatos Prognosticados e Classificados da Análise da Região 1 no Domínio N-terminal da LSU de Rubisco circundando os Resíduos-alvos TYR20, GLU60 e ASN123. Gp e Gm sobrescritos denotam o resíduo de *Galdieria partita* e *Griffithsia monolis*, respectivamente.

N°#	Fig[a]	Mutante	Classificação	Região Afetada
1a	9, 10	Y25W/D51I^{Gp}	3	Adicionar uma interação hidrofóbica nova entre a extremidade C-terminal de αB e βA. βA está perto de Y20 na sequência.
1b	9, 10	Y25W/D51V^{Gm}	3	Adicionar uma interação hidrofóbica nova entre a extremidade C-terminal de αB e βA. βA está perto de Y20 na sequência.
2	-	A59G/G64A	6	Trocar mutação. Interação entre αB e a cadeia longa que conecta αB a βC. O grupo metila trocado estaa espacialmente perto de Y20 e adjacente a E60.
3	-	P49D/D51I	10	Alterar a interação na alça curta que conecta αB e βB.
4	9, 10	T23G/K81R	4	Interação entre βA e βC é quebrada. Afeta o posicionamento de Y20.
5a	10	G54A^{Gp}/C84A/I87V	5	Introduzir uma interação hidrofóbica forte entre αB e βC.
5b	10	G54S^{Gm}/C84A/I87V	5	Introduzir uma interação hidrofóbica forte entre αB e βC.
6a	9	T23G/Y25W/D51I^{Gp}/K81R	1	Mutante #1a adiciona uma interação hidrofóbica entre βA e αB, enquanto Mutante #4 quebra a interação de βA com βC. Estes dois conjuntos de mutações podem ter um efeito grande em Y20.
6b	9	T23G/Y25W/D51V^{Gm}/K81R	1	Mutante #1b adiciona uma interação hidrofóbica entre βA e aB, enquanto Mutante #4 quebra a interação de βA com βC. Estes dois conjuntos de mutações podem ter um efeito grande em Y20.
7a	10	Y25W/D51I^{Gp}/G54A^{Gp}/C84A/I87V	2	Efeito cumulativo dos Mutantes #1a e #5a em E60.
7b	10	Y25W/D51I^{Gp}/G54S^{Gm}/C84A/I87V	2	Efeito cumulativo dos Mutantes #1a e #5b em E60.
7c	10	Y25W/D51V^{Gm}/G54A^{Gp}/C84A/I87V	2	Efeito cumulativo dos Mutantes #1b e #5a em E60.
7d	10	Y25W/D51V^{Gm}/G54S^{Gm}	2	Efeito cumulativo dos Mutantes #1b e #5b

		/C84A/I87V		em E60.
8	11	V12I/M297G/V300T	7	Interação hidrofóbica entre as duas LSUs é quebrada (resíduos 297 e 300 são da LSU vizinha que contém o complexo de Mg do sítio ativo sendo considerado). Afeta N123.
9	11	L36I^{Gm}/I116L/F140L^{Gp}	8	Forma uma região hidrofóbica grande envolvendo as extremidades dos dois fitas β adjacentes (βB e βE) e βC. Poderia alterar o orientação/posicionamento de N123.
10	11	L36I^{Gm}/I116L / V121I/F140L^{Gp} / M297G/V300T	3	Poderiam agir junto em α - C e têm um efeito cumulativo em N123.
11		T114A/T118A/T271V/V 121I	2	Interação polar de 114 e 118 com 271 na LSU para é quebrada. 271 forma uma interação hidrofóbica nova com 121. Impacta N123
12	17	K18I/T23G	5	Interação polar entre cadeias laterais de T23 e K18 é quebrada. Impacta Y20.
13a	9, 10	Y25W/D51I^{Gp}/inserção A21	6	(1a + A21) Inserção de A21 move K21 para longe do resíduo 51 e forma uma interação hidrofóbica nova com I51.
14	17	KLTYY-(21-25)-AKMGYW	4	A forma de uma bobina adjacente a Y20 é alterada, afeta a orientação de Y20. Envolve uma inserção (M); vide Figura 6.
15	17	K18I/KLTYY-(21-25)-AKMGYW/K81R	1	A alteração da forma de bobina próxima a Y20 é associada com alterações a sua interação com 18 (perda de interação polar) e 81 (alteração em comprimento) e pode ter um efeito coordenado em Y20.
16	-	A15S/K18I/T68V/L407I	9	Possível interação entre os resíduos 18 e 68 que poderiam ligar a cauda N-terminal ao resto do domínio; resíduo conservado 69 interage com 407 da LSU par; S15 pode formar ligação de H forte com grupos carbonila da cadeia principal de 408 e 409 da LSU par em algas vermelhas. Alveja Y20.

[a] Figuras 9-11 e 17 mostram os sítios de mutação prognosticados.

EXEMPLO 4

IDENTIFICAÇÃO, AGRUPAMENTO E CLASSIFICAÇÃO DOS RESÍDUOS CANDIDATOS, E PREDIÇÃO DOS MUTANTES CANDIDATOS USANDO MÉTODO DE ENXERTO FILOGENÉTICO ESTENDIDO

[000227] Como acima mencionado e como mostrado nos Fluxogramas 1 e 3, o método de enxerto filogenético estendido fornece um meio para eficazmente utilizar conhecimento do "mapa" de função para mutações, desenvolvido da experiência adquirida dos ciclos de aplicação do método de núcleo de predição e testagem dos Mutantes

Candidatos. Uso do método estendido é ilustrado neste exemplo com referência aos resultados iniciais para os Mutantes Candidatos da Região 1 prognosticados (Exemplo 3) dados na Tabela 5 e debatidos em mais detalhe no Exemplo 7.

[000228] O método estendido foi primeiro usado para interpretar os resultados do método de núcleo. Isto resultou na identificação de um modelo novo para agrupamento de CRs, ACRs e DCRs já identificados (vide CMs na Tabela 3) que foi com base nas Sub-regiões como pontos de acesso mutáveis que focalizado em redes de interações como no método de núcleo. As três Sub-regiões (1A, 1B e 1C) exibiram a propriedade de serem prognosticadas influenciar as propriedades de um dos três Resíduos-alvos preferencialmente ligados à Região 1: Sub-região IA para TYR20, IB para ASN123 e 1C para GLU60. Este "atracamento" dos três TRs às Sub-regiões da Região 1 é mostrado graficamente nas Figuras 15. As Sub-regiões são mostradas graficamente em uma orientação alternativa na Figura 17. As Sub-regiões são confinadas completamente no domínio N-terminal, com exceção dos fragmentos do domínio C-terminal da LSU adjacente que faz parte da Sub-região 1B. Identificação das Sub-regiões proveu a base para identificação de CRs e DCRs adicionais, usando inspeção visual e outras análises previamente descritas para identificar CRs e DCRs usando o método de núcleo que pode ser preferencialmente agrupados com CRs e DCRs já identificados pelo método de núcleo para formar Mutantes Candidatos novos (vide Tabela 4). CRs e DCRs adicionais assim identificadas eram as posições 19, 68, 88 e 104, e 86, 117 e 138, respectivamente. Alternativamente, a identificação das Sub-regiões proveem uma base nova para reagrupar os CRs e DCRs preferenciais já identificados pelo método de núcleo para formar Mutantes Candidatos novos.

[000229] O uso do método estendido, e a base para recrutar os CRs

e DCRs adicional particulares, é ilustrado abaixo com referência aos exemplos para refinamento das atividade dos mutantes prognosticados pelo método de núcleo e inicialmente testados (Tabela 5). Os CMs prognosticados novos são mostrados na Tabela 4.

[000230] TABELA 4. Mutantes Candidatos prognosticados usando o método de enxerto filogenético estendido para refinar os Mutantes Candidatos inicialmente prognosticados mais promissores (Tabela 3). Também listadas são mutações adicionais de componente de resíduo simples e duplo destes Mutantes Candidatos inicialmente prognosticados mais promissores.

N°.	Fig	Mutação	Classificação	Comentários
17-1A	17	T23G/K18I/T68V/K81R	2	Predição com base no enxerto filogenético refinado na sub-região 1A.
18a-1A	17	T23G/K81R/P104E^{Gp}	5	
18b-1A	17	T23G/K81R/P104D^{Gm}	5	
19-1A	17	T23G/D19P/K81R	4	
20-1B	16, 17	I116L/L117T/V121I/ I138M/F140L	3	Predição com base no enxerto filogenético refinado na sub-região 1B. Alveja resíduo N123.
21-1A,B	17	T23G/K81R/V121I/M297G/ V300T	1	Combinação de dois CMs prognosticados que mostraram melhoria geral máxima na eficiência da enzima em estudo de avaliação cinética.
22-1A	9,10	T23G		Componentes de resíduo simples do Mutante #4.
23-1A	9,10	K81R		
24	I1	V121I/M297G		Componente de resíduo duplo do Mutante #8.
25	I1	M297G		Componentes de resíduo simples do Mutante #8.
26-1B	I1	V121I		

[000231] O primeiro exemplo abaixo focaliza no refinamento do melhor mutante, Mutante #4 (T23G/K81R), da testagem de primeiro ciclo, e na sub-região IA. Como mostrado na Tabela 3, Mutante #4 é um componente com Mutante #1a (Y25W/D51I^{Gp}) de Mutante #6a (T23G/Y25W/D51I^{Gp}/K81R). Porém, embora o Mutante #6a mostrou especificidade aumentada (8,5%, vide Tabela 5), não mostrou nenhuma melhoria de eficiência, e geral é inferior ao Mutante #4 que mostrou tanto especificidade aumentada (10%) como eficiência (8%). Re-

ferência aos resultados para o Mutante #1a indicou ser um mutante relativamente pobre com especificidade aumentada de apenas 3,4% e nenhuma alteração na eficiência. Desse modo, concluiu-se que o agrupamento das mutações do Mutante #4 com as do Mutante #1a foi ineficaz, como o efeito líquido não foi uma melhoria cumulativa mas do contrário os dois grupos de mutações interferiram aparentemente de modo desvantajoso. Como a Sub-região IA contém as posições de resíduo das mutações no Mutante #4, deduziu-se que acoplamento das mutações do Mutante #4 com mutações de outros CRs e/ou DCRs na sub-região IA, pode prover uma estratégia melhor para melhorar também nas propriedades deste mutante "guia".

[000232] Exame da sequência e dados estruturais por procedimentos de análise visual e outros já descritos, especialmente uma análise de dados de covariação espécie-específica para CRs e DCRs identificados na Sub-região IA, identificou três CRs adicionais, 19, 68 e 104, que podem influenciar o resíduo de TR TYR20 ou 81 (componente de #Mutante #4). Consequentemente, os CMs novos a seguir, mostrando as alterações indicadas nas interações, foram prognosticados como melhorias potenciais no Mutante #4.

[000233] Em relação ao Mutante #17-1A (#4 + K18I/T68V), as mutações de K181 e T68V são prognosticadas resultar em uma interação alterada entre a cauda N-terminal e a região entre a hélice βB e B. Em combinação com as mutações no Mutante #4 (T23G e K81R) este é esperado ter um impacto aumentado em TR TYR20.

[000234] Em relação ao Mutante #18a(b)-1A (#4 + $P104E^{Gp}(D^{Gm})$), a mutação de CR 104 (que é também um ACR), isto é, $P104E^{Gp}$ ou $P104D^{Gm}$, é esperada resultar em uma interação nova com a cadeia principal N do resíduo 81. É esperado que as três mutações no CM #18-1A combinado teria um impacto aumentado em TR TYR20.

[000235] Em relação ao Mutante #19-1A (#4 + D19P), o resíduo 19

forma uma ligação em ponte de sal com o resíduo 21 (K ou R) em cianobactérias e plantas, que está ausente em algas vermelhas onde o resíduo 19 é P. É esperado que a mutação D19P influênciará para o alcance de TR TYR20 no sítio ativo, e que em combinação com as mutações do Mutante #4, esta produziria um impacto aumentado em TR TYR20.

[000236] Estas predições são mostradas graficamente nas Figuras 15 e 17.

[000237] O segundo exemplo abaixo refere-se à identificação de uma área mutacional nova principal (sub-região 1B) que por sua vez levou à predição inicial do Mutante #20-1B (V116L/L117T/V121I/I138M/F140L). Este mutante tem Mutante #8 (V121I/M297G/V300T), identificado pelo método de núcleo (Tabela 3), como um componente. Como mostrado na Tabela 5, testagem inicial do Mutante #8 mostrou melhoria na especificidade (5,6%) que provendo uma base para mais caracterização. Seguindo identificação da Sub-região 1B, como descrita, exame de variações de resíduo próximo do resíduo 121 (CR adjacente a TR N123; vide Figura 16A) levou à descoberta de uma rede de interações entre a hélice C e β-folha E. Esta rede pode ser vista mais claramente na Figura 16B. Alterações nesta rede de interação podem afetar a orientação de TR N123. As interações envolvem resíduos 116, 117, 120, 121, 124, 133, 135, 138 e 140. Destes, resíduo 121 é um CR e resíduos 120, 124, 133 e 135 são conservados em todos os grupos principais dos organismos (vide Figura 5). Resíduos 116, 117, 138 e 140 são DCRs, isto é, diferem entre os três grupos de organismos principais, algas vermelhas, cianobactérias e plantas floríferas (Tabela 2). Foi observado que alterações no CR e DCRs interagindo nesta região espacial (isto é, 21, e 116, 117, 138 e 140) parecer ser compensatórias nos grupos de organismos principais. Desse modo, como um teste inicial do efeito de tais

alterações na orientação de N123, Mutante #20-1B (V116L/L117T/V1211/I138M/ F140L) foi prognosticado (Tabela 4).

EXEMPLO 5

GERAÇÃO DE RUBISCOS MUTANTES E TRIAGEM DOS MESMOS

[000238] Com base nos Mutantes Candidatos prognosticados classificados nas Tabelas 3 e 4, estudos de prova-de-princípios foram conduzidos mutando os resíduos específicos algais vermelhos na Região 1 de PCC7942 de *Synechococcus* sp., em grupos de mutações múltiplas. Inicialmente quatorze dos mutantes da LSU prognosticados detalhados no Exemplo 3 e listados na Tabela 3 (Mutantes #1a, #lb, #4, #5a, #5b, #6a, #7a, #7b, #7d, #8, #9, #10, #12 e #14, onde "a" e "b" referem-se às variantes onde o CR para *G. partita* ou o ACR para *G. monolis* foram transplantados, respectivamente), foram engenheirados mutando o gene de *rbc*LS de PCC7942 de *Synechococcus* sp. com o kit de multimutagênese QuickChange (Stratagene) usando iniciadores apropriados. Além disso, vários mutantes com mutações simples (#22-1A, #23-1A, #25, #26-1B) ou duplas (#24), que são componentes dos mutantes prognosticados (#4, #8, #21-1A,B), foram engenheirados como controles.

[000239] A sequência genômica para a sequência de *rbc*L-*rbc*S (óperon) de PCC7942 de *Synechococcus* sp. do 'tipo selvagem' é mostrada em SEQ ID N°. 22. Esta sequência é igual à para PCC6301 de *Synechococcus* sp. Na sequência de genoma nativa, a sequência de codificação de *rbc*LS lê ATG CCC (codificando Met depois Pro), mas em todas as sequências de *rbc*LS clonadas é ATG GCC (codificando Met depois Ala). Esta substituição de nucleotídeo foi introduzida para codificar para um único sítio de restrição (*Nco*I) usado para clonagem do gene. Uma mutação silenciosa no segundo último códon (Arg) na sequência de *rbc*SS (CGA para CGC) foi também introduzida para propósitos de clonagem. A sequência transladada para a subunidade

grande do tipo selvagem é mostrada em SEQ ID N°. 24 e a sequência transladada para a subunidade pequena do tipo selvagem é mostrada em SEQ ID N°. 23. As sequências de nucleotídeo e de proteína para mutantes são mostradas através da SEQ ID NOS na Tabela 5, por exemplo, SEQ ID NOS: 25 e 26, respectivamente, para o Mutante #1a.

[000240] Após consideração destes resultados iniciais e desenvolvimento do método de enxerto filogenético estendido, um quatro mais distante dos mutantes de LSU prognosticados detalhados no Exemplo 4 e listados na Tabela 4 (Mutante #17-1A, #18-1A, #19-1A, #21-1A5B) foi engenheirado pelo mesmo método. Tudos os mutantes de controle (#22-1A, #23-1A, #24, #25, #26-1B) mencionados acima são componentes destes quatro mutantes e são relevantes, em particular, para os Mutantes #4 e #21-1A,B. As sequências de nucleotídeo e de proteína para estes mutantes são mostradas pelas SEQ ID NOS na Tabela 5, por exemplo, SEQ ID NOS: 53 e 54, respectivamente, para o Mutante #17-1A.

[000241] Os genes de *rbc*LS mutados foram sequênciados antes de clonar de volta para o segundo plasmídeo de expressão que codificou para a LSU mutada e a SSU nativa. As Rubiscos mutante foram expressadas e purificadas usando um procedimento descrito no Exemplo 6. Estes experimentos iniciais no dezoito Mutantes de PCC7942 de *Synechococcus* sp. e cinco mutantes de controle mostraram expressão boa em *E. coli* das Rubiscos de mutantes ativos (isto é, corretamente enovelados e montados hexadecaméricos), com especificidade e constantes cinéticas comparáveis, ou melhor que do tipo selvagem, como descrito no Exemplo 7. O modo em que teste experimental e procedimentos de optimização, detalhados nos Exemplos 6-11 é integrado com a predição e etapas de triagem *in silico* é mostrado no Fluxograma 1. Fluxograma 1 também mostra como os resultados experimentais podem ser realimentados no procedimento de predição para

refinar as predições de proteínas mutantes, seguido por mais ciclos de triagem *in silico* e triagem experimental e avaliação da melhoria na propriedade funcional para otimizá-lo.

EXEMPLO 6

EXPRESSÃO E PURIFICAÇÃO DE RUBISCOS MUTANTE

[000242] Embora o *E. coli* seja o hospedeiro microbiano amplamente usado para expressar o DNA recombinante e proteínas, obter a forma funcional da Rubisco de organismos eucarióticos (definida aqui como proteína de "forma I") é mais complexo. Como a holoenzima de Rubisco de forma I é um hexadecâmero feito de 8 subunidades grandes (LSUs) e 8 subunidades pequenas (SSUs), ela requer corretamente enovelar e montar a enzima para chaperonas apropriadas corretamente. Porém, convenientemente quando o óperon que codifica para os genes de Rubisco (*rbc*LS e *rbc*SS) de PCC7942 de *Synechococcus* sp. é expressado em *E. coli* as subunidades de LSU e de SSU são sintetizadas abundantemente. Consequentemente, PCC7942 de *Synechococcus* sp. foi usado aqui como a enzima de L_8S_8 modelo para testagem inicial das predições de Rubisco, e *E. coli* foi usado como a escolha natural do hospedeiro de expressão para produzir as Rubiscos mutantes.

[000243] Porém, apenas cerca de 1 a 5% das LSUs expressas são corretamente enoveladas e montadas em forma funcional com a quantidade de Rubisco funcional acumulando em ~1 a 3% (p/p) da proteína solúvel de *E. coli*. Purificação da Rubisco funcional através de abordagens tradicionais é um processo laborioso e demorado que pode levar 3 dias.

[000244] O uso de marcadores de afinidade de 6x Histidina (H_6) é uma alternativa atrativa que poderia economizar esforço substancial e tempo para purificação da enzima. Mas os experimentos mostraram que a fusão de marcadores de H_6 em quaisquer términos da LSU ou SSU

da Rubisco de forma I pode comprometer a atividade catalítica. Para superar estas dificuldades, um sistema recentemente adaptado (Baker et al, 2005, os conteúdos inteiros deste são aqui incorporados por referência) foi usado para simplificar e acelerar a purificação da Rubisco de PCC7942 de *Synechococcus* sp. expressa em *E. coli*. Este sistema envolveu a construção de um vetor de plasmídeo único (com base em pACYC) incorporando fundir na estrutura a sequência de codificação para uma sequência de ubiquitina (Ub) marcada com H_6 (H_6Ub) na extremidade de 5' de *rbc*SS. O *rbc*LS do tipo selvagem no plasmídeo pTrcSynLS (que contém o óperon de *rbc*L-*rbc*S de PCC7942; Emlyn-Jones et al, 2006) foi substituído com o *rbc*LS mutado (*rbc*LS *) e depois cotransformado em *E. coli* com o plasmídeo com base em pACYC que codifica para SSU do tipo selvagem marcado com H_6Ub (H_6Ub-SSU). Quando expressão da subunidade da Rubisco foi induzida com IPTG, todos os três peptídeos da subunidade da Rubisco foram produzidos (isto é, LSU, SSU e H_6UbSSU). Alguns foram montados em hexadecâmeros de Rubisco funcional compostos de 8xLSU núcleos octaméricos e razões diferentes de SSU (no máximo 8) e H_6UbSSU. Rubiscos com um ou mais marcadores de H_6 são facilmente purificados de outras proteínas de *E. coli* usando cromatografia de afinidade de metal imobilizada (IMAC) e a sequência de H_6Ub depois clivada com uma protease H_6Ub-específica que, junto com peptídeos de H_6Ub desmontados, pode ser removida por IMAC. Usando este método, Rubiscos purificadas foram isoladas do *E. coli* em cerca de 1 hora.

[000245] Dezoito das Rubiscos mutantes e cinco mutantes de controle especificados no Exemplo 5 e identificados nas Tabelas 3 ou 4, como também tipo selvagem, foram expressados e purificados usando o procedimento acima.

EXEMPLO 7

ENSAIO CINÉTICO *IN VITRO* DE RUBISCOS MUTANTES

[000246] Proteínas de Rubisco purificadas usando o método descrito no Exemplo 6 foram usadas para medir as constantes de Michaelis para CO_2 (Kc) e as taxas de carboxilação de substrato saturado (V_c^{max}) usando ensaios de fixação de $^{14}CO_2$ a 25°C, pH 8 de acordo com o método descrito em Andrews (1988), os conteúdos inteiros deste são incorporados aqui por referência.

[000247] A enzima purificada foi pré-incubada a 25°C por 30 min em tampão contendo 20 mM $MgCl_2$ e 25 mM $NaHCO_3$, e medições de Kc foram executadas em frascos de cintilação revestidos com septo pulverizado com nitrogênio. As reações foram iniciadas adicionando 10 µl da enzima purificada a 0,5 ml de tampão de ensaio equilibrado com N_2 (100 mM EPPS-NaOH, 20 mM $MgCl_2$, 0,8 mM ribulose-P_2, 0,1 mg/ml anidrase carbônica) contendo concentrações variadas de $NaH_{14}CO_3$.

[000248] As constantes de Michaelis foram determinadas ajustando os dados à Equação de Michaelis-Menten. Quantificação do conteúdo de Rubisco nos ensaios foi medida usando o ensaio de ligação [2-^{14}C] carboxiarabinitol-P_2 (^{14}2CABP) descrito por Ruuska et al. (1998) e Whitney e Andrews (2001). A taxa de giro de carboxilação saturada de substrato (k_c^{cat}) foi calculada dividindo a atividade de carboxilase máxima extrapolada (V_c^{max}) pela concentração dos sítios ativos de Rubisco no ensaio. As preparações de Rubiscos purificadas foram também usadas para medir a especificidade de CO_2/O_2 ($S_{c/o}$) em pH 8.3 como descrito em Kane et al. (1994).

[000249] Um sumário dos resultados obtido é apresentado na Tabela 5. Para especificidade, os resultados médios são dados como % de alteração comparados com o tipo selvagem: um valor positivo representa uma melhoria. Para cinética, os valores e % de alteração comparados com o tipo selvagem são dados; uma % de alteração positiva para a taxa catalítica k_c^{cat} e a eficiência catalítica k_c^{cat}/k_c representa uma melhoria, enquanto uma % de alteração negativa para a constan-

te de Michaelis k_c representa uma melhoria.

[000250] TABELA 5. Resultados de especificidade e cinética para Rubiscos mutantes de PCC7942 de *Synechococcus* sp. Os números na coluna "SEQ ID" correspondem às sequências nas listagens de sequência gerada por computador, n.m. representa não medido no momento do depósito.

Mutante # (resíduo(s) modificado(s))[c]	**SEQ ID[e]**	**$S_{c/o}$**	**Cinética**		
		Média[d] (% de alteração do tipo selvagem)	**K^c_{cat} (s^{-1}) (% de alteração do tipo selvagem**	**K_c (µM) (% de alteração do tipo selvagem)**	**K^c_{cat}/K_c (S^{-1}-mM^{-1}) (% de alteração do tipo selvagem)**
Tipo Selvagem	22-24	41,6 ± 0,7	13,2 ± 0,2[a] 13,0 ± 0,2[b]	203 ± 10[a] 197 ± 10[b]	65[a] 66[b]
#1a (Y25W, D51I^{Gp})	25,26	(3,4%)	11,5 ± 0,2[a] (-13%)	176 ± 6[a] (-13%)	65 (0%)
#1b (Y25W, D51V^{Gm})	27,28	(-1,1 %)			
#4 (T23G, K81R)	29,30	(10,0%)	14,1 ± 0,2[b] (8%)	198 ± 8[b] (0%)	71 (8%)
#5a (G54A^{Gp}, C84A, I87V)	31,32	(-6%)	11,3 ± 0,3[a] (-14%)	182 ± 12[a] (-10%)	62 (4%)
#5b (G54S^{Gm}, C84A, I87V)	33,34	(1,9%)	10,6 ± 0,2[a] (-20%)	167 ± 8[a] (-18%)	63 (3%)
#6a (T23G, Y25W, D51I^{Gp}, K81R)	35,36	(8,5%)	12,2 ± 0,2[b] (-6%)	185 ± 9[b] (-6%)	66 (0%)
#7a (Y25W, D51I^{Gp}, G54A^{Gp}, C84A, I87V)	37,38	(8,8%)	12,4 ± 0,2[b] (-5%)	195 ± 11[b] (-1%)	64 (-3%)
#7b (Y25W, D51I^{Gp}, G54S^{Gm}, C84A, I87V)	39,40	(1,4%)	n.m.	n.m.	n.m.
#7d (Y25W, D51V^{Gm}, G54A^{Gp}, C84A, I87V)	41,42	(0,9%)	n.m.	n.m.	n.m.
#8 (V121I, M297G, V300T)	43,44	(5,6%)	n.m.	n.m.	n.m.
#9 (L36I, I116L, F140L)	45,46	(3,6%)	n.m.	n.m.	n.m.
#10 (L36I, I116L, V121I, F140L, M297G, V300T)	47,48	(2,9%)	10,8 ± 0,4[b] (-17%)	325 ± 27[b] (65%)	33 (-50%)
#12 (K18I, T23G)	49,50	(1%)	n.m.	n.m.	n.m.
#14 (alça AKMGYW)	51,52	(1,7%)	7,1 ± 0,4[b] (-45%)	260 ± 36[b] (32%)	27 (-59%)
#17-1A (T23G, K18I, T68V, K81R)	53,54	(-0,6%)	n.m.	n.m.	n.m.
#18-1A (T23G, K81R, P104E)	55,56	(-0,5%)	n.m.	n.m.	n.m.
#19-1A (T23G, D19P, K81R)	57,58	(-2,0%)	n.m.	n.m.	n.m.
#21-1A,B (T23G, K81R, V121I, M297G, V300T)	59,60	(-0,9%)	n.m.	n.m.	n.m.
#22-1A (T23G)	61,62	(5,0%)	n.m.	n.m.	n.m.
#23-1A (K81R)	63,64	(7,0%)	n.m.	n.m.	n.m.
#24 (V121I, M297G)	65,66	n.m.	n.m.	n.m.	n.m.
#25 (M297G)	67,68	n.m.	n.m.	n.m.	n.m.
#26-1B (V121I)	69,70	n.m.	n.m.	n.m.	n.m.

a,b - refere-se às medições replicadas nas amostras de Rubisco diferentes feitas em dias diferentes.

c - onde dado, especifica mutações estão nas Sub-regiões 1A, 1A,B ou 1B.

d - média das medições replicadas nas amostras de Rubisco diferentes feitas em dias diferentes.

e - Números de ID da Sequência correspondem aqueles no depósito da sequência.

[000251] Dos dezoito mutantes e dois mutantes simples (controles) para os quais medições de especificidade foram feitas, todos mostraram especificidade comparável com o tipo selvagem, isto é, nenhuma foi prejudicada significativamente, enquanto cinco mutantes e os dois controles mostraram melhorias de 5% ou melhor. De interesse, quatro destes incluem as mutações T23G e/ou K81R. Também, de interesse é que mutantes com variantes de ACR (#1a e #lb para posição 51, #5a e #5b para posição 54, e #7a, #7b e #7d para posições 51 e 54) mostraram diferenças significativas (na ordem de 8% para as variantes dos Mutantes #5 e #7) demonstrando a sensibilidade de especificidade para estas alterações que espacialmente estão relativamente longe do sítio ativo (vide Figura 10). A melhoria no valor de $S_{c/o}$ de 5,6% no Mutante #8 foi notável como as mutações estavam em uma parte diferente do domínio N-terminal (vide Figura 15) que a maioria do conjunto inicial dos mutantes prognosticados pelo método de núcleo.

[000252] Dos 9 mutantes para os quais k_c^{cat} foram ensaiados, a maioria mostrou valores ligeira a significativamente inferiores (isto é, mais fracos) comparados com tipo selvagem com a exceção, Mutante #4 que mostrou uma melhoria significativa de 8%. Os 9 mutantes correspondentes ensaiados para K_c exibiram uma faixa de valores, com 4 mostrando ligação de CO_2 moderadamente melhorada (#1a, #5a, #5b, #6a), 2 mostrando pequena alteração (#4, #7a) e 2 mostrando prejuízo significativo (#10, #14). Os valores de eficiência catalítica geral

(k_c^{cat}/k_c) similarmente mostram uma faixa de pouca a significativa melhoria (#4, #5a, #5b), para pequena alteração (#1a, #6a, #7a) para prejuízo significativo (#10, #12) quando comparados com o tipo selvagem. É notável que em tabaco (vide Exemplo 11), o Mutante #23-1A mostrou um valor de k_c^{cat} significativamente mais fraco e também eficiência catalítica geral, enquanto que o Mutante #4 mostrou melhorias significativas em todas as três medidas cinéticas.

[000253] Dos resultados do conjunto inicial de mutantes prognosticados pelo método de núcleo (#1 - #14), os mutantes superiores em termos de desempenho geral foram Mutantes #4 e #6a que mostraram melhorias em especificidade e eficiência cinética de 10% e 8% e 8,5% e 0%, respectivamente. As propriedades do mutante mais complexo (#6a), que compreendia as mutações em #4 e #1a, foram no geral inferiores ao Mutante #4 (o melhor mutante na lista atual). Esta observação sugere que outras mutações possam ser mais vantajosamente agrupadas com as para o Mutante #4 que as do Mutante #1a. Como debatido no Exemplo 4, esta observação incitou o desenvolvimento do conceito de Sub-regiões e do método estendido, que levou às primeiras predições dos Mutantes #17-1A, #18-1A e #19-1A, como também um mutante mais complexo incluindo CRs de Sub-região 1B também (#21-1A3B). Os resultados de especificidade para estas predições de método estendido mostraram pequena alteração comparada ao tipo selvagem.

EXEMPLO 8

EVOLUÇÃO DIRECIONADA DE MUTANTES DE *Synechococccus* EM *E. Coli*

[000254] Os mutantes enxertados filogenéticos representam uma estratégia "direcionada" por explorar as regiões espaciais da sequência não naturalmente amostradas. Porém qualquer aumento na atividade destes mutatantes pode ser prejudicado devido a algumas áreas

de "sequência fraca se encaixa", visto que elas podem não ser otimizadas para a estrutura da Rubisco do hospedeiro. Embora o método de enxerto filogenético estendido forneça uma estratégia racional *in silico* que pode ser usada para otimizar os mutantes "guias", incluindo aliviar os conflitos estéricos por recrutamento de SvRs e outros resíduos de variante de ocorrência natural que possam ser identificados como complementares para as mutações "guias", uma opção alternativa para minimizar os efeitos destes conflitos pode se usar um método de evolução direcionado experimental para os otimizar, isto é, usar estas Rubiscos parcialmente otimizadas como pontos de partida. Estes mutantes também proveem neles mesmos um ponto de partida novo para evolução direcionada visto que eles têm potencial diferente para explorar espaço da sequência comparados com o tipo selvagem.

[000255] Como detalhado no Exemplo 6, diferente de todas as outras Rubiscos de Forma I (isto é, hexadecaméricas) de organismos eucarióticos, Rubisco de PCC7942 de *Synechococcus* pode montar-se corretamente em *E. coli* e foi escolhida para um procedimento de triagem de mutante. Usando os abordagens descritos nos Exemplos 5-7, predições de Mutante de Candidato (que podem incluir grupos de mutações correlatadas, e mutações independentes em regiões estruturais diferentes circundando os Resíduos-alvos) podem ser triadas inicialmente em *E. coli* para confirmar que eles são ativos e obter constantes de cinética *in vitro* para comparação um ao outro e ao tipo selvagem. Mutantes com até 10-12 mutações podem ser produzidos habitualmente usando a tecnologia atual. Como detalhado nos Exemplos 5-7, mutantes simples selecionados podem ser feitos para testar a hipótese geral subjacente do método de enxerto filogenético que os mutantes simples são prováveis de ser fracamente ativos/inativos, e que pelo menos duas mutações correlatadas são necessárias para produzir uma enzima aceitavelmente ativa.

[000256] Um sistema adequado para evolução direcionada de Rubisco em *E. coli* foi relatado recentemente (Mueller-Cajar et al., 2007). Este usa uma cepa de *E. coli* engenheirada, MM1, cujo crescimento pode ser feito dependente da expressão funcional de Rubisco, quando coexpressada com fosforribulocinase (PRK). Glicólise em MM1 foi bloqueada mediante deleção do gene de gliceraldeído 3-fosfato desidrogenase (gapA) e um desvio do desvio metabólico compreendendo uma PRK de *Synechococcus* e Rubisco de *Rodospirillum rubruma* foi introduzida. Como resultado, MM1 é dependente da expressão de Rubisco funcional para metabolizar o produto de catálise de PRK, ribulose-1,5-bisfosfato, que é tóxico para *E. coli*. Este método geral pode ser usado para avaliar se as Rubiscos com atividade significativamente intensificada podem ser evoluídas mais eficientemente a partir de formas inativadas dos mutantes enxertados com PCC7942 de *Synechococcus* sp. mais promissores detalhados no Exemplo 7 e Tabela 5. Para este propósito, bibliotecas fortuitamente mutageneizadas (feitas usando os abordagens relatados por Mueller-Cajar et al., 2007) destes genes inativados podem ser transformadas em células MM1 cultivadas sob condições seletivas diferentes (por exemplo variando as pressões de CO_2/O_2 de cultivo, alterando a extensão da produção de PRK). Colônias expressando variantes de Rubisco evoluídas com capacidade física melhorada (isto é, aquelas que sobrevivem à triagem) podem ser isoladas, sequênciadas e a cinética das Rubiscos mutadas purificadas caracterizada como detalhada na Exemplo 7.

EXEMPLO 9

TESTE DA COMPETÊNCIA BIOQUÍMICA E FISIOLÓGICA DOS MUTANTES DE *Synechococcus in vivo*

[000257] Os testes funcionais *in vitro* no Exemplo 7 identificaram vários mutantes de *Synechococcus* com atividade de Rubisco melhorada, e que também, necessariamente, foram, desse modo, corretamen-

te enovelados e montados quando expressos do sistema de expressão de *E. coli* e purificados como descritos nos Exemplo 6. Na estratégia de reengenharia da Rubisco, *Synechococcus* foi usado como o hospedeiro inicial mais conveniente para o teste experimental das predições de mutantes para identificar candidatos "guias". Devido aos recentes avanços em criar Rubisco mutante em uma planta modelo, tabaco, usando transformação de plastídio, como descrito no Exemplo 10, no trabalho descrito aqui um Mutante Candidato (#4) identificado como um mutante de *Synechococcus* promissor foi testado diretamente na planta florífera de teste (tabaco) sem empreender a etapa intermediária, mostrada no Fluxograma 1, de confirmar que ele pode ser reaclimado em seu hospedeiro nativo (*Synechococcus*) sem detrimento para seu fenótipo. Este método melhorado permitiu avaliação da montagem e cinética das Rubiscos mutantes em tabaco dentro de 7-9 semanas o que é mais rápido que requerido para se obter e cultivar transformantes de cianobactérias. Onde o último propósito de desenvolvimento de mutante for para criar Rubisco melhorada em plantas, esta estratégia seria preferida. Porém, se o último propósito for para criar Rubisco melhorada em cianobactérias, então é preferido executar esta etapa de teste fisiológico em *Synechococcus* para avaliar se as mutações cataliticamente benéficas poderiam introduzir problemas de incompatibilidade tais como os que perturbam o enovelamento-montagem produtivos por complexos de chaperona de cianobactérias ou outras proteínas de interação tais como as envolvidas na formação de carboxissoma. Um método para executar estes testes em *Synechococcus* é descrito abaixo.

[000258] Em PCC7942 de *Synechococcus* sp., os genes de Rubisco (*rbc*LS e *rbc*SS) são codificados por um óperon simples no cromossomo (do qual há tipicamente 5 cópias de cromossomo por célula) e, análogo ao *E. coli*, esta cepa de cianobactérias é naturalmente compe-

tente e pode ser geneticamente transformada ou através de modificações alvejadas para seu cromossomo (por exemplo deleção de gene, substituição de gene) através de recombinação homóloga ou por retenção estável dos vetores ponte do plasmídeo dentro de suas células.

[000259] Cepas mutantes de PCC7942 de *Synechococcus* foram desenvolvidas (vide Emlyn-Jones et al, 2006 e Price et al., 1993 para exemplos). Uma cepa mutente de PCC7942 de *Synechococcus* em que o óperon de *rbc*LS-*rbc*SS cromossômico é deletado (isto é, uma cepa de 7942Δ*rbc*LS) pode ser usada para facilitar a reintrodução dos genes mutantes de Rubisco *rbc*LS-*rbc*SS. Como PCC7942 de *Synechococcus* sp. não pode desenvolver-se heterotroficamente (isto é, em uma fonte de carbono externa) e requer uma Rubisco para crescimento, cepas de 7942Δ*rbc*LS podem ser geradas: (1) introduzindo um segundo gene de Rubisco (por exemplo *rbc*M codificando para o homodímero de Rubisco da Forma II estruturalmente diferente (L_2) da bactéria *Rodospirillum rubrum* ou a Rubisco de L_8S_8 de PCC7942 nativa) em um vetor ponte de plasmídeo em PCC7942 de *Synechococcus* sp. depois (2) recombinar homologamente em um gene de resistência a antibióticos km^R para substituir o óperon *rbc*LS-*rbc*SS em cada cópia do cromossomo (isto é, assim a mutação pode ser segregada completamente). Células de PCC7942 de *Synechococcus* transformadas com *rbc*M expresso em um vetor ponte podem ser subsequentemente transformadas com outro plasmídeo para homologamente substituir a região de codificação de *rbc*LS-*rbc*SS cromossômica com um gene de km^R. Sob isolamento dos transformantes de Δ*rbc*LS-*rbc*SS::km^R completamente segregados (isto é, todos os cromossomos têm *rbc*LS-*rbc*SS substituídos com km^R) as células de PCC7942Δ*rbc*LS podem ser usadas para homologamente reintroduzir os genes mutados de *rbc*LS* e *rbc*SS e as células curadas do vetor ponte.

[000260] Usando técnicas estabelecidas (por exemplo vide Emlyn-

Jones et al, 2006), o fenótipo das células transformadas pode depois ser compreensivamente caracterizado de modo bioquímico, por exemplo, avaliando se as subunidades das Rubiscos mutadas são facilmente enoveladas e corretamente montadas, e fisiologicamente, por exemplo, avaliando se há diferenças na capacidade fotossintética, partição de carbono inorgânico ou taxa de crescimento.

EXEMPLO 10

GERAÇÃO DE TRANSFORMANTES DE PLASTOMA DE RUBISCO EM TABACO

[000261] Uma estratégia para reengenheirar Rubisco que é aplicável a Rubiscos de planta mais alta é através da transformação de plastoma usando genes *rbc*L* mutados. O plastoma de tabaco é facilmente transformável (Andrews e Whitney, 2003) e foi usado como um modelo para conduzir testes de prova-de-princípio para a transformação de outras plantas.

[000262] O Mutante #4 de Rubisco de *Synechococcus* mais promissor foi usado como o caso de teste inicial. Os mutantes simples de componente, T23G (#22-1A Mutante), e K81R (Mutante #23-1A) foram também testados. O nucleotídeo e a proteína SEQ ID NOS para Mutantes de tabaco #4 e #23-1A são dados na Tabela 6. As SEQ ID NOS para Mutante #22-1A são 77 e 78. Como estes resíduos são Resíduos Candidatos (vide Tabela 2) estas mesmas mutações foram usadas em tabaco como em *Synechococcus*.

[000263] Linhagens transplastômicas de tabaco estão disponíveis que permitem triagem mais rápida da cinética das Rubiscos mutadas de tabaco que os abordagens de transformação de cloroplasto prolongados tradicionais em que os genes de *rbc*L nativos no plastoma são substituídos com versões mutadas (Andrews e Whitney, 2003 (supra)) usando a técnica de transformação biolística descrita em Svab e Maliga (1993). Em uma recente melhoria para o método para transformar

genes de *rbc*L mutados ou estranhos de volta no plastoma de tabaco, o gene de *rbc*L nativo foi substituído com o gene de *rbc*M de *Rodospirillum rubrum* e o gene marcador selecionável ctadA (codificando para resistência à espectinomicina) e depois o gene de *aadA* removido para produzir linhagens transplastômicas de tabaco-*rubrum* livre de marcador (Δ*aadA*). Estas linhagens foram geradas biolisticamente transformando o plastoma de tabaco do tipo selvagem (*Nicotiana tabacum* L. cv Petit Havana [*N,N*]) com plasmídeo p^{cm}trLA (número de acesso do Genbank AY827488). O gene de *aadA* nos transformantes é flanqueado por sítios *loxP* de 34 bp que permitem sua excisão através de recombinação de CRE-lox. Para excisar *aadA*, as folhas de uma linhagem transformada com p^{cm}trLA foram biolisticamente bombardeadas com o plasmídeo de expressão de CRE pKO27 (Corneille et al, 2001). As folhas bombardeadas foram dissecadas (~0,5 cm^2) e propagadas em meio seletivo de canamicina (sais de Murashige-Skoog solidificados com ágar contendo 3% (p/v) de sucrose, 15 μg ml^{-1} de canamicina e hormônios (Svab e Maliga, 1993). As primeiras plantinhas para emergir das seções das folhas bombardeadas alvejadas foram transferidas para o meio MS (meio seletivo sem canamicina ou hormônios) e perda de *aadA* confirmada por análises de PCR rotineiras.

[000264] As linhagens de tabaco-*rubrum* de Δ*aadA* permitem reuso do gene marcador de *aadA* para subsequentemente transformar seu plastoma, tal como retransformar de volta em variantes de *rbc*L* de tabaco mutado para substituir *rbc*M. A eficiência da transformação de substituir *rbc*M no tabaco-*rubrum* de Δ*aadA* com genes de *rbc*L* variantes é 3 a 10 vezes mais alta que transformar tabaco do tipo selvagem, e é imune a eventos de recombinação indesejados que podem ocorrer quando transformar genes de *rbc*L* em plastomas de tabaco do tipo selvagem. O método permitiu a produção de transformantes contendo apenas cópias de plastoma transformado de *rbc*L* (isto é,

transformantes homoplásticos) dentro de 6 a 8 semanas como as cópias de plastoma contendo o gene *rbc*M foram rapidamente eliminadas. Como a Rubisco de *R. rubrum* é um homodímero pequeno de LSUs (L_2, ~100 kDa), linhagens transformadas de *rbc*L* que produzem as Rubiscos de L_8S_8 de maior forma I (~520 kDa) podem ser identificadas separando a proteína da folha solúvel por eletroforese em gel de poliacrilamida não-desnaturante como descrito em Whitney e Sharwood (2007) e homoplasmicidade medida pela ausência de Rubisco de L_2.

[000265] Este sistema é potencialmente adaptável para outras plantas onde transformação de plastídio foi relatada. A transformação genética dos plastídios em uma variedade de plantas diferentes foi relatada, por exemplo, em Koop et al. (1997), os conteúdos inteiros deste são incorporados aqui por referência.

[000266] Mutações para Mutantes #4, #22-1A e #23-1A foram feitos para a sequência de codificação de *rbc*L de plastoma de tabaco do tipo selvagem (*Nicotiana tabacum*) usando o kit de multimutagênese QuickChange (Stratagene) usando iniciadores apropriados, e introduzidos no plasmídeo de transformação de plastoma de tabaco, pLEV1, onde a seleção dos transformantes é facilitada pela incorporação de um gene *aadA* menos promotor a jusante de *rbc*L (Whitney et al., 1999). Os plasmídeos de transformação derivados de pLEV1 codificando as cópias de *rbc*L* de tabaco mutageneizado com as sequências de nucleotídeo codificando para os Mutantes #4, #22-1A e #23-1A, como também o tipo selvagem, foram biolisticamente transformados em uma linhagem de tabaco-*rubrum* de *ΔaadA* e plantinhas resistentes à espectinomicina selecionadas como descrito (Svab e Maliga, 1993). Nenhuma plantinha foi isolada para Mutante #22-1A. Transformantes onde o *rbc*M tinha sido substituído com os genes *rbc*L* ou *rbc*L foram identificados pela produção de Rubisco de L_8S_8 usando eletrofo-

rese em gel de poliacrilamida não-desnaturante (vide Figura 20).

EXEMPLO 11

CARACTERIZAÇÃO BIOQUÍMICA E FISIOLÓGICA DE TRANSFORMANTES DE RUBISCO DE TABACO

[000267] Os abordagens a seguir foram usados para extrair e purificar Rubisco expressadas em células de tabaco homoplásmicas, transplastômicas e do tipo selvagem para levar a cabo as análises cinéticas.

[000268] Ensaios de fixação de $^{14}CO_2$ radiomarcados foram usados para medir a taxa de giro do substrato saturado (k_c^{cat}) e constante de Michaelis para CO_2 a 0% (k_c) ou 21% O_2 (k_c^{ar}) usando extrato de proteína de folha solúvel. Discos foliares (~1 cm^2) foram tirados durante o fotoperíodo e extraídos em gelo usando homogeneizadores de vidro (Wheaton, USA) em 0,8 ml de tampão de extração livre de CO_2 (50 mM Bicina-NaOH, pH 8,0, 1 mM EDTA, 2 mM DTT, 1% (v/v) coquetel inibidor da protease de planta (Sigma-Aldrich) e 1% (w/v) PVPP). A amostra foi centrifugada (36.000 g, 5 min, 40ºC) e a proteína solúvel incubada (ativada) com $NaHCO_3$ e $MgCl_2$ (15 mM cada) por 15 min e usada para medir o conteúdo de Rubisco em alíquotas duplicadas incubadas com 40 µM de ^{14}2-CABP e a quantidade de ^{14}2-CABP ligado à Rubisco restabelecida por filtração em gel (Ruuska et al, 1998), ou usado para medir k_c e k_c^{ar} em 25ºC, pH 8,0 usando ensaios de fixação de $^{14}CO_2$ (Andrews 1988; Whitney e Sharwood, 2007). Para confirmar que as amostras usadas eram homoplásmicas (isto é, apenas contêm cópias de plastoma transformadas com os genes *rbc*L* e nenhuma com *rbc*M) a proteína foi também separada em géis de poliacrilamida não-desnaturantes (Figura 20). Como para os controles do tipo selvagem, apenas Rubisco de L_8S_8 mutada e nenhuma Rubisco de L_2 foi expressado no Mutante #4 e transformantes de #23-1A.

[000269] Os ensaios cinéticos foram iniciados adicionando extrato de

proteína solúvel ativada em frascos de cintilação revestidos com septo contendo ou N2 (para k_c) ou tampão de ensaio equilibrado com ar livre de CO_2 (para k_c^{ar}) (100 mM Bicina-NaOH, 15 mM $MgCl_2$, 0,6 mM ribulose-P_2, 0,1 mg-ml^{-1} anidrase carbônica) contendo 0 a 90 µM $^{14}CO_2$. Ribulose-P_2 foi sintetizada de acordo com (Kane et al, 1998). Os ensaios foram parados após 1 min com 0,5 volume de 25% (v/v) de ácido fórmico, as reações secadas a 80°C e o resíduo dissolvido em água, dois volumes de cintilante foram adicionados, submetido a vórtice, e ^{14}C medido por contagem de cintilação. k_c e k_c^{ar} foram calculados do diagrama de Michaelis-Menten da taxa de carboxilação versus [CO_2].

[000270] As medições de k_c^{cat} foram calculadas usando ensaios de fixação de $^{14}CO_2$ comparáveis contendo 15 mM $NaH^{14}CO_3$ e dividindo a taxa de carboxilação de substrato saturado, V_c^{max}, pela concentração dos sítios ativos de Rubisco medidos por ligação de ^{14}2-CABP (vide acima).

[000271] Medições de especificidade foram feitas usando Rubisco purificada. Proteína de folha solúvel foi como descrita acima e 0,2 ml cromatografado através de uma coluna Superdex 200HR 10/30 equilibrada com tampão de especificidade (30 mM Trietanolamina pH 8,3, 30 mM acetato de Mg) usando um sistema explorador ÄKTA (APBiotech). As três frações de pico (0,3 ml) contendo Rubisco de L_8S_8 foram agrupadas (~100-150 pmol de sítios de subunidade L) e usadas para medir especificidade de CO_2/O_2 a 25ºC como descrito (Kane et al, 1994) após equilibrar com uma atmosfera contendo 500 ppm de CO_2 em O_2 controlado usando três bombas de mistura de gás de precisão Wösthoff.

[000272] Um sumário dos resultados obtidos é apresentado nas Tabelas 6 e 7. Para especificidade, resultados médios são dados como % de alteração comparado com o tipo selvagem: um valor positivo representa uma melhoria. Para cinética, os valores e % de alteração

comparados com o tipo selvagem são dados; um % de alteração positivo para a taxa catalítica k^c_{cat} e a eficiência catalítica k^c_{cat}/k_c representa uma melhoria, enquanto um % de alteração negativo para a constante de Michaelis k_c representa uma melhoria.

TABELA 6. Resultados da especificidade e cinética medidos no ar (O_2 21%) para Rubiscos mutantes de Tabaco.

Mutante # (resíduo(s) modificado(s))	**Seq ID[a]**	**$S_{c/o}$**	**Cinética**		
		Média (% de alteração do tipo selvagem)	**K^c_{cat} (s^{-1}) (% de alteração do tipo selvagem**	**K_c^{ar} (µM) (% de alteração do tipo selvagem)**	**K^c_{cat}/K_c^{ar} (S^{-1}-mM^{-1}) (% de alteração do tipo selvagem)**
tipo selvagem	71,72	81,7 ± 2,1	3,59 ± 0,02	20,8 ± 1,4	173
#4 (T23G, K81R)	73,74	79,3 ± 1,4 (-3%)	3,84 ± 0,05 (7%)	18,2 ± 0,6 (-13%)	209 (21%)
#23-1A (K81R)	75,76	80,7 ± 1,9 (-1%)	2,46 ± 0,17 (-31%)	17,8 ± 1,0 (-14%)	138 (-20%)

a - Números de ID da Sequência correspondem àqueles na listagem de sequência; o primeiro número é a SEQ ID NO do nucleotídeo e o segundo número é a SEQ ID NO da proteína.

[000273] Os resultados na Tabela 6 para medições em ar (21% O_2) mostram melhorias significativas no tipo selvagem em valores k_{cat} e k_c^{ar} para Mutante #4 de 7% e -13%, respectivamente, produzindo uma melhoria geral na eficiência catalítica de 21%. Porém, embora o Mutante simples #23-1A mostre uma melhoria similar em k_c^{ar} (-14%), ele tem uma taxa significativamente prejudicada (-31%) produzindo uma diminuição geral na eficiência catalítica de 20%. Os resultados da cinética (ensaios em triplicata) foram obtidos usando extrato de proteína de folha solúvel de dois (tipo selvagem, #23-1A) ou três (#4) transformantes diferentes. Os Rubsicos isolados replicados deram resultados similares, contribuindo para erros pequenos das medições.

[000274] Os resultados da especificidade para ambos mutantes mostraram pequena alteração comparada com o tipo selvagem. Estes valores de medições quadruplicadas foram obtidos usando Rubiscos purificadas de dois transformantes do tipo selvagem, dois de Mutante #4

e dois de Mutante #23-1A.

[000275] Consideração dos resultados de especificidade e de cinética juntos indica que a melhoria na eficiência de carboxilação para Mutante #4 foram equiparada com uma melhoria comparável em eficiência de oxigenação. Como um teste inicial de se alterações na eficiência de oxigenação poderiam ser devido às alterações em k^o_{cat} ou K_o, k_c foi medida em 0% O_2. Estes resultados, dados na Tabela 7, mostram que ambos os mutantes têm valores de k_c mais fracos que o tipo selvagem, 18% e 9% mais alto para Mutantes #4 e #23-1A, respectivamente. Isto indica os valores de K_o para os mutantes são melhores que o tipo selvagem, ou seja, eles são menos inibidos por O_2. Valores aproximados para K_o, mostrados na última coluna, foram calculados de acordo com Whitney et al. (1999). Como estas medições de K_o são com base nas medições de k_c em apenas duas concentrações de O_2 (isto é, 0 e 21% (v/v) em nitrogênio), elas deveriam ser consideradas como estimativas apenas. Porém, assumindo que elas são seguros então pode-se concluir que o k^o_{cat} para os mutantes deve ser melhorado para explicar a melhoria na eficiência de oxigenação sugerida pelos resultados de especificidade.

[000276] TABELA 7. Resultados de especificidade e de cinética medidos em 0% O_2 para Rubiscos mutantes de Tabaco.

Mutante # (resíduo(s) modificado(s))	**Cinética**			
	K^c_{cat} (s^{-1}) (% de alteração do tipo selvagem	**K_c (µM) (% de alteração do tipo selvagem)**	**K^c_{cat}/K_c (S^{-1} mM^{-1}) (% de alteração do tipo selvagem)**	**$K_i(O_2)$ (µM)**
tipo selvagem	3,59 ± 0,02	12,7 ± 0,1	283	395
#4 (T23G, K81R)	3,84 ± 0,05 (7%)	15,0 ± 0,7 (18%)	256 (-10%)	1157
#23-1A(K81R)	2,46 ± 0,17 (-31%)	13,9 ± 0,6 (9%)	177 (-37%)	884

[000277] Em resumo, os resultados mostram que o Mutante #4 mantem suas propriedades superiores em tabaco, embora elas sejam expressadas mais como melhorias na eficiência catalítica que como me-

lhorias em termos gerais iguais em especificidade e eficiência, como é o caso para o Mutante #4 de *Synechococcus*.

[000278] Linhagens transplastômicas homoplásmicas e controles de tabaco do tipo selvagem podem ser crescidas até a maturidade em terra em câmaras de crescimento de ambiente controlado ou casas de vidro e testes fisiológicos padrão depois empreendidos. Estes testes são feitos apenas para as Rubiscos mutantes mais promissoras mostrando melhorias na especificidade e/ou cinética no tipo selvagem e incluem determinação da capacidade fotossintética, taxa de crescimento e bioquímica da folha, incluindo nível de expressão de Rubisco, metabólito, níveis de carbono e nitrogênio usando metodologias estabelecidas (Whitney et al, 1999; Whitney e Andrews, 2001; Whitney et al., 2001).

REFERÊNCIAS

[000279] Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 25, 3389-3402.

[000280] Andrews TJ. (1988) Catalysis by cianobacterial ribulose-bisphosphate carboxylase large subunits in the complete absence of small subunits. *J. Biol. Chem.* 263, 12213-12220.

[000281] Andrews TJ, Whitney SM. (2003) Manipulating ribulose bisphosphate carboxylase/oxygenase in the cloroplasts of higher plants. *Arch. Biochem. Biophys.* 414, 159-169.

[000282] Baker RT, Catanzariti AM, Karunasekara Y, Soboleva TA, Sharwood R, Whitney S, Board PG. (2005) Using deubiquitylating enzymes as research tools. *Methods Enzymol.* 398, 540-554.

[000283] Case DA, Darden TA, Cheatham III, TE, Simmerling CL, Wang J, Duke RE, Luo R, Merz KM, Pearlman DA, Crowley M, Walker RC, Zhang W, Wang B, Hayik S, Roitberg A, Seabra G,

Wong KF, Paesani F, Wu X, Brozell S, Tsui V, Gohlke H, Yang L, Tan C, Mongan J, Hornak V, Cui G, Beroza P, Mathews DH, Schafmeister C, Ross WS, Kollman PA. (2006) *AMBER 9,* University of California, San Francisco.

[000284] Ciniglia C, Yoon HS, Pollio A, Pinto G, Bhattacarya D. (2004) Hidden biodiversity of the extremophilic *Cyanidiales* red algae. *Mol. Ecol.* 13, 1827-1838.

[000285] Corneille S, Lutz K, Svab Z, Maliga P. (2001) Efficient elimination of selectable marker genes from the plastid genome by the CRE-lox site-specific recombination system. *Plant J.* 27, 171-178.

[000286] Cummins PL. (1996) Molecular Orbital Programs for Simulations (MOPS), Australian National University, Canberra.

[000287] Cummins PL, Gready JE. (1997) A coupled semiempirical molecular orbital and molecular mechanical model (QM/MM) for organic molecules in aqueous solution. *Comput. Chem.* 18, 1496-1512.

[000288] Cummins PL, Gready JE. (1998) A molecular dynamics and free energy perturbation (MD/FEP) study of the hydride-ion transfer step in dihydrofolate reductase using a combined quantum and molecular mechanical (QM/MM) model. J. Comput. Chem. 19, 977-988.

[000289] Cummins PL, Gready JE. (1999) Coupled semiempirical quantum mechanics and molecular mechanics model (QM/MM) calculations on the aqueous solvation energies of ionised molecules. J. *Comput. Chem.* 20, 1028-1038.

[000290] Cummins PL, Gready JE. (2003) Computational methods for the study of enzymic reaction mechanisms II: An overlapping mechanically embedded method for hybrid semiempirical-QM/MM calculations. *THEOCHEM* 632, 245-255.

[000291] Cummins PL, Gready JE. (2005) Computational methods for the study of enzymic reaction mechanisms III: a perturbation plus QM/MM approach for calculating relative free energies of protonation. J

Comput. Chem. 26, 561-568.

[000292] Cummins PL, Rostov I, Gready JE. (2007) Calculation of a complete enzymic reaction surface: reaction and activation free energies for hydride-ion transfer in dihydrofolate reductase. *J. Chem. Theor. Comput.* 3, 1203-1211.

[000293] Emlyn-Jones D, Woodger FJ, Price GP, Whitney SM. (2006) RbcX can function as a rubisco chaperonin, but is non-essential in PCC7942 de *Synechococcus*. *Plant Cell Physiol.* 47, 1630-1640.

[000294] Evans JR, Austin RB. (1986) The specific activity of ribulose-1,5-bisphosphate carboxylase in relation to genotype in wheat. *Planta* 167, 344-350.

[000295] Fersht A. (1998) Structure and mechanism in protein science: guide to enzyme catalysis and protein folding. W. H. Freeman & Co., 1998.

[000296] Frey PA, Hegeman A. (2007) Enzymatic reaction mechanisms. Oxford University Press USA, 2007.

[000297] Frisch MJ, (80 coautores) e Pople JA. (2004) *Gaussian 03,* Revision C.02, Gaussian Inc., Wallingford, CT.

[000298] Galmés J, Flexas J, Keys AJ, Cifre J, Mitchell RAC, Madgwick PJ, Haslam RP, Medrano H, Parry MAJ. (2005) Rubisco specificity factor tends to be larger in plant species from drier habitats and in species with persistent leaves. *Plant Cell Environ.* 28, 571-579.

[000299] Gready JE, Rostov I, Cummins PL. (2006) Simulations of enzyme reaction mechanisms in active sites: accounting for an environment which is much more than a solvent perturbation. Em "Modelling Molecular Structure and Reactivity in Biological Systems", K.J. Naidoo, M. Hann, J. Gao, M. Field e J, Brady, eds, Royal Society of Chemistry, Londres, páginas 101-118.

[000300] Hall TA. (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucl.*

Acids Symp. Ser. 41, 95.

[000301] Kane HJ, Viil J, Entsch B, Paul BK, Morell MK, Andrews TJ. (1994) An improved method for measuring the CO2/02 specificity of ribulosebisphosphate carboxylaseoxygenase. *Aust. J. Plant Physiol.* 21, 449-461.

[000302] Kane HJ, Wilkin JM, Portis AR, Andrews TJ. (1998) Potent inhibition of ribulosebisphosphate carboxylase by an oxidized impurity in ribulose-1,5-bisphosphate. *Plant Physiol.* 117, 1059-1069.

[000303] Koop H-U, Herz S, Golds TJ, Nickelsen J. (2007) The genetic transformation of plastids. Topic in Current Genetics. DOI 10.1007/4735 2007 0225/Publicado em-linha: 15 de maio de 2007.

[000304] Mueller-Cajar O, Morell M, Whitney SM. (2007) Directed evolution of Rubisco in *Escherichi coli* reveals a specificity-determining hydrogen bond in the Form II enzyme. Biochemistry, em impressão, setembro de 2007.

[000305] Parry MA, Andralojc PJ, Mitchell RA, Madgwick PJ, Keys AJ. (2003) Manipulation of Rubisco: the amount, activity, function and regulation. *J. Exptl Bot.* 54, 1321-1333.

[000306] Parry MAJ, Madgwick PJ, Carvalho JFC, Andralojc PJ. (2007) Prospects from increasing photosynthesis by overcoming the limitations of Rubisco. *J. Ag. Sci.* 145, 3143.

[000307] Price GD, Howitt SM, Harrison K, Badger MR. (1993) Analysis of a genomic DNA region from the cianobacterium *Synechococcus* sp. strain PCC7942 involved in carboxysome assembly and function. J. *Bacteriol.* 175, 2871-2879.

[000308] Ruuska S, Andrews TJ, Badger MR, Hudson GS, Laisk A, Price GD, von Caemmerer S. (1998). The interplay between limiting processes in C-3 photosynthesis studied by rapid-response gas exchange using transgenic tobacco impaired in photosynthesis. *Aust. J. Plant Physiol.* 25, 859-870.

[000309] Svab Z, Maliga P. (1993) High-frequency plastid transformation in tobacco by selection for a chimeric *aadA* gene. *Proc. Natl Acad. Sci. USA* 90, 913-917.

[000310] Spreitzer RJ, Salvucci ME. (2002) Rubisco: structure, regulatory interactions, and possibilities for a better enzyme. *Annu. Rev. Plant. Biol.* 53, 449-475.

[000311] Thompson JD, Higgins DG, Gibson TJ. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* 22, 4673-4680.

[000312] Whitney SM, von Caemmerer S, Hudson GS, Andrews TJ. (1999) Directed mutation of the Rubisco large subunit of tobacco influences photorespiration and growth. *Plant Physiol.* 121, 579-588.

[000313] Whitney SM, Andrews TJ. (2001) Plastoma-encoded bacterial ribulose-1,5-bisphosphate carboxylase/oxygenase (Rubisco) supports photosynthesis and growth in tobacco. *Proc. Natl Acad. Sci. USA* 98, 14738-14743.

[000314] Whitney SM, Baldet P, Hudson GS, Andrews TJ. (2001). Form I Rubiscos from non-green algae are expressed abundantly but not assembled in tobacco cloroplasts. *Plant J.* 26, 535-547.

[000315] Whitney SM, Sharwood RE. (2007) Linked Rubisco subunits can assemble into functional oligomers without impeding catalytic performance. J. *Biol. Chem.* 282, 3809-3818.

Listagem de Sequência

<110> The Australian National University

<120> Processo para geração de proteína e usos das mesmas

<130> 819973

<150> AU 2006905622
<151> 2006-10-10

<160> 76

<170> patente em versão 3.2

<210> 1
<211> 475
<212> PRT
<213> Cyanophora paradoxa

<400> 1

```
Met Ser Ser Gln Ala Arg Thr Gln Thr Arg Ala Gly Phe Lys Ala Gly
1               5                  10                  15
Val Lys Asp Tyr Arg Leu Thr Tyr Tyr Thr Pro Glu Tyr Thr Pro Lys
            20                  25                  30
Glu Thr Asp Ile Leu Ala Ala Phe Arg Met Thr Pro Gln Pro Gly Val
        35                  40                  45
Pro Pro Glu Glu Cys Ala Ala Ala Val Ala Ala Glu Ser Ser Thr Gly
    50                  55                  60
Thr Trp Thr Thr Val Trp Thr Asp Gly Leu Thr Ser Leu Asp Arg Tyr
65                  70                  75                  80
Lys Gly Arg Ser Tyr Gly Phe Glu Pro Val Pro Gly Glu Glu Asn Gln
                85                  90                  95
Tyr Ile Cys Tyr Val Ala Tyr Pro Leu Asp Leu Phe Glu Glu Gly Ser
            100                 105                 110
Val Thr Asn Met Leu Thr Ser Ile Val Gly Asn Val Phe Gly Phe Lys
        115                 120                 125
Ala Leu Arg Ala Leu Arg Leu Glu Asp Leu Arg Ile Pro Val Gly Tyr
    130                 135                 140
Ser Lys Thr Phe Gln Gly Pro Pro His Gly Ile Thr Val Glu Arg Asp
145                 150                 155                 160
Lys Leu Asn Lys Tyr Gly Arg Ala Leu Leu Gly Cys Thr Ile Lys Pro
                165                 170                 175
Lys Leu Gly Leu Ser Ala Lys Asn Tyr Gly Arg Ala Val Tyr Glu Cys
            180                 185                 190
Leu Arg Gly Gly Leu Asp Phe Thr Lys Asp Asp Glu Asn Val Asn Ser
        195                 200                 205
Gln Pro Phe Met Arg Trp Arg Asp Arg Phe Leu Tyr Val Met Asp Ala
    210                 215                 220
Ile Lys Lys Ser Gln Ala Glu Thr Gly Glu Ile Lys Gly His Tyr Leu
225                 230                 235                 240
Asn Ala Thr Ala Pro Thr Thr Glu Glu Met Ile Lys Arg Ala Glu Phe
                245                 250                 255
Ala Ala Glu Leu Asp Ala Pro Ile Ile Met His Asp Tyr Ile Thr Ala
            260                 265                 270
Gly Phe Thr Ser Asn Thr Thr Leu Ala Arg Trp Cys Arg Asp Asn Gly
        275                 280                 285
Pro Leu Leu His Ile His Arg Ala Met His Ala Val Ile Asp Arg Gln
```

```
    290                 295                 300
Lys Asn His Gly Ile His Phe Arg Val Leu Ala Lys Thr Leu Arg Met
305                 310                 315                 320
Ser Gly Gly Asp His Leu His Ser Gly Thr Val Val Gly Lys Leu Glu
                325                 330                 335
Gly Asp Arg Ala Gly Thr Leu Gly Phe Val Asp Leu Met Arg Asp Asp
            340                 345                 350
His Ile Glu Gln Asp Arg Ser Arg Gly Ile Phe Phe Thr Gln Asp Trp
        355                 360                 365
Ala Ser Met Pro Gly Val Met Pro Val Ala Ser Gly Gly Ile His Ile
    370                 375                 380
Trp His Met Pro Ala Leu Val Asp Ile Phe Gly Asp Asp Ser Cys Leu
385                 390                 395                 400
Gln Phe Gly Gly Gly Thr Leu Gly His Pro Trp Gly Asn Ala Pro Gly
                405                 410                 415
Ala Val Ala Asn Arg Val Ala Leu Glu Ala Cys Val Gln Ala Arg Asn
            420                 425                 430
Glu Gly Arg Asn Leu Ala Arg Glu Gly Asn Glu Ile Ile Arg Glu Ala
        435                 440                 445
Ala Arg Phe Ser Pro Glu Leu Ala Ala Ala Cys Glu Val Trp Lys Glu
    450                 455                 460
Ile Lys Phe Glu Phe Glu Thr Ile Asp Thr Ile
465                 470                 475
```

<210> 2
<211> 488
<212> PRT
<213> Artificial

<220>
<223> 50% da sequência de consenso para rhodophyta (9 espécies) Rubisco LSU

<220> Região de interesse biológico
<221>
<222> (2)..(2)
<223> resíduo é um de A, C, D, G, N, P, S, T ou V

<220>
<221> MISC_FEATURE
<222> (3)..
<223> resíduo é um de C, D, E, H, K, N, Q, R, S ou T

<220>
<221> Região de interesse biológico
<222> (6)..(6)
<223> resíduo é um de C, D, E, H, K, N, Q, R, S ou T

<220>
<221> Região de interesse biológico
<222> (93)..(93)
<223> resíduo é um de A, G ou S

<220>
<221> Região de interesse biológico
<222> (97)..(97)
<223> resíduo é um de S ou T

<220>
<221> Região de interesse biológico

<222> (159)..(159)
<223> resíduo é um de I, L ou V

<220>
<221> Região de interesse biológico
<222> (219)..(219)
<223> resíduo é um de H, K ou R

<220>
<221> Região de interesse biológico
<222> (233)..(233)
<223> resíduo é um de C, D, E, H, K, N, Q, R, S ou T

<220>
<221> Região de interesse biológico
<222> (236)..(236)
<223> resíduo é um de S ou T

<220>
<221> Região de interesse biológico
<222> (239)..(239)
<223> resíduo é um de I, L ou V

<220>
<221> Região de interesse biológico
<222> (253)..(253)
<223> resíduo é um de D ou E

<220>
<221> Região de interesse biológico
<222> (269)..(269)
<223> resíduo é um de I, L ou V

<220>
<221> Região de interesse biológico
<222> (309)..(309)
<223> resíduo é um de A, C, D, G, N, P, S, T ou V

<220>
<221> Região de interesse biológico
<222> (343)..(343)
<223> resíduo é um de I, L ou V

<220>
<221> Região de interesse biológico
<222> (346)..(346)
<223> resíduo é um de H, K ou R

<220>
<221> Região de interesse biológico
<222> (355)..(355)
<223> resíduo é um de S ou T

<220>
<221> Região de interesse biológico
<222> (358)..(358)
<223> resíduo é um de D ou E

<220>
<221> Região de interesse biológico
<222> (359)..(359)

<223> resíduo é um de I, L ou V

<400> 2

```
Met Xaa Xaa Ser Val Xaa Glu Arg Thr Arg Ile Lys Asn Glu Arg Tyr
1               5                  10                  15
Glu Ser Gly Val Ile Pro Tyr Ala Lys Met Gly Tyr Trp Asp Pro Asp
            20                  25                  30
Tyr Val Val Lys Asp Thr Asp Val Leu Ala Leu Phe Arg Val Thr Pro
        35                  40                  45
Gln Pro Gly Val Asp Pro Val Glu Ala Ser Ala Ala Val Ala Gly Glu
    50                  55                  60
Ser Ser Thr Ala Thr Trp Thr Val Val Trp Thr Asp Leu Leu Thr Ala
65                  70                  75                  80
Cys Asp Leu Tyr Arg Ala Lys Ala Tyr Lys Val Asp Xaa Val Pro Asn
                85                  90                  95
Xaa Ser Asp Gln Tyr Phe Ala Tyr Ile Ala Tyr Asp Ile Asp Leu Phe
            100                 105                 110
Glu Glu Gly Ser Ile Ala Asn Leu Thr Ala Ser Ile Ile Gly Asn Val
        115                 120                 125
Phe Gly Phe Lys Ala Val Lys Ala Leu Arg Leu Glu Asp Met Arg Leu
    130                 135                 140
Pro Val Ala Tyr Leu Lys Thr Phe Gln Gly Pro Ala Thr Gly Xaa Ile
145                 150                 155                 160
Val Glu Arg Glu Arg Met Asp Lys Phe Gly Arg Pro Phe Leu Gly Ala
                165                 170                 175
Thr Val Lys Pro Lys Leu Gly Leu Ser Gly Lys Asn Tyr Gly Arg Val
            180                 185                 190
Val Tyr Glu Gly Leu Lys Gly Gly Leu Asp Phe Leu Lys Asp Asp Glu
        195                 200                 205
Asn Ile Asn Ser Gln Pro Phe Met Arg Trp Xaa Glu Arg Tyr Leu Tyr
    210                 215                 220
Ser Met Glu Gly Val Asn Arg Ala Xaa Ala Ala Xaa Gly Glu Xaa Lys
225                 230                 235                 240
Gly His Tyr Leu Asn Val Thr Ala Ala Thr Met Glu Xaa Met Tyr Glu
                245                 250                 255
Arg Ala Glu Phe Ala Lys Glu Leu Gly Ser Val Ile Xaa Met Ile Asp
            260                 265                 270
Leu Val Ile Gly Tyr Thr Ala Ile Gln Thr Met Ala Ile Trp Ala Arg
        275                 280                 285
Lys Asn Asp Met Ile Leu His Leu His Arg Ala Gly Asn Ser Thr Tyr
    290                 295                 300
Ser Arg Gln Lys Xaa His Gly Met Asn Phe Arg Val Ile Cys Lys Trp
305                 310                 315                 320
Met Arg Met Ala Gly Val Asp His Ile His Ala Gly Thr Val Val Gly
                325                 330                 335
Lys Leu Glu Gly Asp Pro Xaa Met Ile Xaa Gly Phe Tyr Asn Thr Leu
            340                 345                 350
Leu Leu Xaa His Leu Xaa Xaa Asn Leu Pro Gln Gly Ile Phe Phe Glu
        355                 360                 365
Gln Asp Trp Ala Ser Leu Arg Lys Val Thr Pro Val Ala Ser Gly Gly
    370                 375                 380
Ile His Cys Gly Gln Met His Gln Leu Leu Asp Tyr Leu Gly Asp Asp
385                 390                 395                 400
Val Val Leu Gln Phe Gly Gly Gly Thr Ile Gly His Pro Asp Gly Ile
                405                 410                 415
Gln Ala Gly Ala Thr Ala Asn Arg Val Ala Leu Glu Ser Met Val Ile
            420                 425                 430
Ala Arg Asn Glu Gly Arg Asp Tyr Val Ala Glu Gly Pro Gln Ile Leu
        435                 440                 445
Arg Asp Ala Ala Lys Thr Cys Gly Pro Leu Gln Thr Ala Leu Asp Leu
```

```
    450                 455                 460
Trp Lys Asp Ile Thr Phe Asn Tyr Thr Ser Thr Asp Thr Ala Asp Phe
465                 470                 475                 480
Val Glu Thr Pro Thr Ala Asn Val
                485
```

<210> 3
<211> 476
<212> PRT
<213> Artificial

<220>
<223> 50% da sequência de consenso para cyanobacteria Rubisco

<220> Região de interesse biológico
<221>
<222> (6)..(6)
<223> resíduo é um de A, C, F, G, H, I, K, L, M, R, T, V, W ou Y

<220> Região de interesse biológico
<221>
<222> (8)..(8)
<223> resíduo é um de S ou T

<220> Região de interesse biológico
<221>
<222> (10)..(10)
<223> resíduo é um de S ou T

<220> Região de interesse biológico
<221>
<222> (12)..(12)
<223> resíduo é um de A, G ou S

<220> Região de interesse biológico
<221>
<222> (15)..(15)
<223> resíduo é um de C, D, E, H, K, N, Q, R, S ou T

<220> Região de interesse biológico
<221>
<222> (43)..(43)
<223> resíduo é um de A, C, F, G, H, I, K, L, M, R, T, V, W ou Y

<220> Região de interesse biológico
<221>
<222> (51)..(51)
<223> resíduo é um de A, C, D, G, N, P, S, T ou V

<220> Região de interesse biológico
<221>
<222> (99)..(99)
<223> resíduo é um de F, H, W ou Y

<220> Região de interesse biológico
<221>
<222> (145)..(145)
<223> resíduo é um de F, H, W ou Y

<220>

<221> Região de interesse biológico
<222> (227)..(227)
<223> resíduo é um de D, E, H, K ou R

<220>
<221> Região de interesse biológico
<222> (246)..(246)
<223> resíduo é um de A, C, D, G, N, P, S, T ou V

<220>
<221> Região de interesse biológico
<222> (252)..(252)
<223> resíduo é um de A, C, F, G, H, I, K, L, M, R, T, V, W ou Y

<220>
<221> Região de interesse biológico
<222> (263)..(263)
<223> resíduo é um de C, D, E, H, K, N, Q, R, S ou T

<220>
<221> Região de interesse biológico
<222> (290)..(290)
<223> resíduo é um de I, L ou V

<220>
<221> Região de interesse biológico
<222> (341)..(341)
<223> resíduo é um de A, G ou S

<220>
<221> Região de interesse biológico
<222> (353)..(353)
<223> resíduo é um de D ou E

<220>
<221> Região de interesse biológico
<222> (354)..(354)
<223> resíduo é um de F, H, W ou Y

<220>
<221> Região de interesse biológico
<222> (357)..(357)
<223> resíduo é um de D, E, H, K ou R

<220>
<221> Região de interesse biológico
<222> (437)..(437)
<223> resíduo é um de D, E, H, K, R

<220>
<221> Região de interesse biológico
<222> (439)..(439)
<223> resíduo é um de A, C, D, E, G, H, K, N, Q, R, S, T

<220>
<221> Região de interesse biológico
<222> (443)..(443)
<223> resíduo é um de A, C, D, G, N, P, S, T ou V

<220>
<221> Região de interesse biológico

<222> (458)..(458)
<223> resíduo é um de I, L ou V

<400> 3

```
Met Ser Tyr Ala Gln Xaa Lys Xaa Gln Xaa Lys Xaa Gly Tyr Xaa Ala
1               5                  10                  15
Gly Val Lys Asp Tyr Arg Leu Thr Tyr Tyr Thr Pro Asp Tyr Thr Pro
            20                  25                  30
Lys Asp Thr Asp Leu Leu Ala Ala Phe Arg Xaa Thr Pro Gln Pro Gly
        35                  40                  45
Val Pro Xaa Glu Glu Ala Gly Ala Ala Val Ala Ala Glu Ser Ser Thr
    50                  55                  60
Gly Thr Trp Thr Thr Val Trp Thr Asp Leu Leu Thr Asp Leu Asp Arg
65                  70                  75                  80
Tyr Lys Gly Arg Cys Tyr Asp Ile Glu Pro Val Pro Gly Glu Asp Asn
                85                  90                  95
Gln Tyr Xaa Ala Phe Ile Ala Tyr Pro Leu Asp Leu Phe Glu Glu Gly
            100                 105                 110
Ser Val Thr Asn Val Leu Thr Ser Ile Val Gly Asn Val Phe Gly Phe
        115                 120                 125
Lys Ala Leu Arg Ala Leu Arg Leu Glu Asp Ile Arg Phe Pro Val Ala
    130                 135                 140
Xaa Ile Lys Thr Phe Gln Gly Pro Pro His Gly Ile Gln Val Glu Arg
145                 150                 155                 160
Asp Lys Leu Asn Lys Tyr Gly Arg Pro Leu Leu Gly Cys Thr Ile Lys
                165                 170                 175
Pro Lys Leu Gly Leu Ser Ala Lys Asn Tyr Gly Arg Ala Val Tyr Glu
            180                 185                 190
Cys Leu Arg Gly Gly Leu Asp Phe Thr Lys Asp Asp Glu Asn Ile Asn
        195                 200                 205
Ser Gln Pro Phe Gln Arg Trp Arg Asp Arg Phe Leu Phe Val Ala Glu
    210                 215                 220
Ala Ile Xaa Lys Ala Gln Ala Glu Thr Gly Glu Ile Lys Gly His Tyr
225                 230                 235                 240
Leu Asn Val Thr Ala Xaa Thr Cys Glu Glu Met Xaa Lys Arg Ala Glu
                245                 250                 255
Phe Ala Lys Glu Leu Gly Xaa Pro Ile Ile Met His Asp Phe Leu Thr
            260                 265                 270
Gly Gly Phe Thr Ala Asn Thr Thr Leu Ala Lys Trp Cys Arg Asp Asn
        275                 280                 285
Gly Xaa Leu Leu His Ile His Arg Ala Met His Ala Val Ile Asp Arg
    290                 295                 300
Gln Lys Asn His Gly Ile His Phe Arg Val Leu Ala Lys Cys Leu Arg
305                 310                 315                 320
Leu Ser Gly Gly Asp His Leu His Thr Gly Thr Val Val Gly Lys Leu
                325                 330                 335
Glu Gly Glu Arg Xaa Ile Thr Met Gly Phe Val Asp Leu Leu Arg Glu
            340                 345                 350
Xaa Xaa Val Glu Xaa Asp Arg Ser Arg Gly Ile Phe Phe Thr Gln Asp
        355                 360                 365
Trp Ala Ser Met Pro Gly Val Met Ala Val Ala Ser Gly Gly Ile His
    370                 375                 380
Val Trp His Met Pro Ala Leu Val Glu Ile Phe Gly Asp Asp Ser Val
385                 390                 395                 400
Leu Gln Phe Gly Gly Gly Thr Leu Gly His Pro Trp Gly Asn Ala Pro
                405                 410                 415
Gly Ala Thr Ala Asn Arg Val Ala Leu Glu Ala Cys Val Gln Ala Arg
            420                 425                 430
Asn Glu Gly Arg Xaa Leu Xaa Arg Glu Gly Xaa Asp Ile Ile Arg Glu
        435                 440                 445
```

```
Ala Ala Lys Trp Ser Pro Glu Leu Ala Xaa Ala Cys Glu Leu Trp Lys
    450                 455                 460
Glu Ile Lys Phe Glu Phe Glu Ala Met Asp Thr Leu
465                 470                 475


<210>  4
<211>  475
<212>  PRT
<213>  Anthoceros formosae

<400>  4

Met Ser Pro Gln Thr Glu Thr Lys Ala Gly Val Gly Phe Lys Ala Gly
1               5                   10                  15
Val Lys Asp Tyr Arg Leu Thr Tyr Tyr Thr Pro Asp Tyr Glu Thr Lys
            20                  25                  30
Asp Thr Asp Ile Leu Ala Ala Phe Arg Met Thr Pro Gln Pro Gly Val
        35                  40                  45
Pro Pro Glu Glu Ala Gly Ala Ala Val Ala Ala Glu Ser Ser Thr Gly
    50                  55                  60
Thr Trp Thr Thr Val Trp Thr Asp Gly Leu Thr Ser Leu Asp Arg Tyr
65                  70                  75                  80
Lys Gly Arg Cys Tyr Asp Ile Glu Pro Val Ala Gly Glu Glu Asn Gln
                85                  90                  95
Tyr Ile Ala Tyr Val Ala Tyr Pro Leu Asp Leu Phe Glu Glu Gly Ser
            100                 105                 110

Val Thr Asn Met Phe Thr Ser Ile Val Gly Asn Val Phe Gly Phe Lys
        115                 120                 125
Ala Leu Arg Ala Leu Arg Leu Glu Asp Leu Arg Ile Pro Pro Ala Tyr
    130                 135                 140
Ser Lys Thr Phe Gln Gly Pro Pro His Gly Ile Gln Val Glu Arg Asp
145                 150                 155                 160
Lys Leu Asn Lys Tyr Gly Arg Pro Leu Leu Gly Cys Thr Ile Lys Pro
                165                 170                 175
Lys Leu Gly Leu Ser Ala Lys Asn Tyr Gly Arg Ala Val Tyr Glu Cys
            180                 185                 190
Leu Arg Gly Gly Leu Asp Phe Thr Lys Asp Asp Glu Asn Val Asn Ser
        195                 200                 205
Gln Pro Phe Met Arg Trp Arg Asp Arg Phe Leu Phe Val Ala Glu Ala
    210                 215                 220
Ile Phe Lys Ser Gln Ala Glu Thr Gly Glu Ile Lys Gly His Tyr Leu
225                 230                 235                 240
Asn Ala Thr Ala Gly Thr Cys Glu Glu Met Met Lys Arg Ala His Phe
                245                 250                 255
Ala Arg Glu Leu Gly Met Pro Ile Val Met His Asp Tyr Leu Thr Gly
            260                 265                 270
Gly Phe Thr Ala Asn Thr Thr Leu Ala Arg Tyr Cys Arg Asp Asn Gly
        275                 280                 285
Leu Leu Leu His Ile His Arg Ala Met His Ala Val Ile Asp Arg Gln
    290                 295                 300
Arg Asn His Gly Ile His Phe Arg Val Leu Ala Lys Ala Leu Arg Met
305                 310                 315                 320
Ser Gly Gly Asp His Ile His Ser Gly Thr Val Val Gly Lys Leu Glu
                325                 330                 335
Gly Glu Arg Glu Val Thr Leu Gly Phe Val Asp Leu Leu Arg Asp Asp
            340                 345                 350
Tyr Ile Glu Lys Asp Arg Ser Arg Gly Ile Tyr Phe Thr Gln Asp Trp
        355                 360                 365
Val Ser Met Pro Gly Val Leu Pro Val Ala Ser Gly Gly Ile His Val
```

```
    370                 375                 380
Trp His Met Ser Ala Leu Thr Glu Ile Phe Gly Asp Asp Ser Val Leu
385                 390                 395                 400
Gln Phe Gly Gly Gly Thr Leu Gly His Pro Trp Gly Asn Ala Pro Gly
                405                 410                 415
Ala Val Ala Asn Arg Val Ala Leu Glu Ala Cys Val Gln Ala Arg Asn
            420                 425                 430
Glu Gly Arg Asp Leu Ala Arg Glu Gly Asn Asp Ile Ile Arg Glu Ala
        435                 440                 445
Ser Lys Trp Ser Pro Glu Leu Ala Ala Ala Cys Glu Val Trp Lys Glu
    450                 455                 460
Ile Lys Phe Val Phe Glu Thr Ile Asp Thr Leu
465                 470                 475

<210>  5
<211>  475
<212>  PRT
<213>  Artificial

<220>
<223>  50% da sequência de consenso para bryophyta (28 espécies)

<400>  5

Met Ser Pro Gln Thr Glu Thr Lys Ala Gly Val Gly Phe Lys Ala Gly
1               5                   10                  15
Val Lys Asp Tyr Arg Leu Thr Tyr Tyr Thr Pro Asp Tyr Gln Thr Lys
            20                  25                  30
Glu Thr Asp Ile Leu Ala Ala Phe Arg Met Thr Pro Gln Pro Gly Val
        35                  40                  45
Pro Ala Glu Glu Ala Gly Ala Ala Val Ala Ala Glu Ser Ser Thr Gly
    50                  55                  60
Thr Trp Thr Thr Val Trp Thr Asp Gly Leu Thr Ser Leu Asp Arg Tyr
65                  70                  75                  80
Lys Gly Arg Cys Tyr Asp Ile Glu Ala Val Pro Gly Glu Glu Asn Gln
                85                  90                  95
Tyr Ile Ala Tyr Val Ala Tyr Pro Leu Asp Leu Phe Glu Glu Gly Ser
            100                 105                 110
Val Thr Asn Leu Phe Thr Ser Ile Val Gly Asn Val Phe Gly Phe Lys
        115                 120                 125
Ala Leu Arg Ala Leu Arg Leu Glu Asp Leu Arg Ile Pro Pro Ala Tyr
    130                 135                 140
Ser Lys Thr Phe Gln Gly Pro Pro His Gly Ile Gln Val Glu Arg Asp
145                 150                 155                 160
Lys Leu Asn Lys Tyr Gly Arg Pro Leu Leu Gly Cys Thr Ile Lys Pro
                165                 170                 175
Lys Leu Gly Leu Ser Ala Lys Asn Tyr Gly Arg Ala Val Tyr Glu Cys
            180                 185                 190
Leu Arg Gly Gly Leu Asp Phe Thr Lys Asp Asp Glu Asn Val Asn Ser
        195                 200                 205
Gln Pro Phe Met Arg Trp Arg Asp Arg Phe Leu Phe Val Ala Glu Ala
    210                 215                 220
Ile Tyr Lys Ser Gln Ala Glu Thr Gly Glu Ile Lys Gly His Tyr Leu
225                 230                 235                 240
Asn Ala Thr Ala Gly Thr Cys Glu Glu Met Leu Lys Arg Ala Gln Phe
                245                 250                 255
Ala Arg Glu Leu Gly Met Pro Ile Val Met His Asp Tyr Leu Thr Gly
            260                 265                 270
Gly Phe Thr Ala Asn Thr Ser Leu Ala His Tyr Cys Arg Asp Asn Gly
        275                 280                 285
Leu Leu Leu His Ile His Arg Ala Met His Ala Val Ile Asp Arg Gln
```

```
    290                 295                 300
Lys Asn His Gly Met His Phe Arg Val Leu Ala Lys Ala Leu Arg Leu
305                 310                 315                 320
Ser Gly Gly Asp His Ile His Ala Gly Thr Val Val Gly Lys Leu Glu
                325                 330                 335
Gly Glu Arg Gln Val Thr Leu Gly Phe Val Asp Leu Leu Arg Asp Asp
            340                 345                 350
Tyr Ile Glu Lys Asp Arg Ser Arg Gly Ile Tyr Phe Thr Gln Asp Trp
        355                 360                 365
Val Ser Leu Pro Gly Val Leu Pro Val Ala Ser Gly Gly Ile His Val
    370                 375                 380
Trp His Met Pro Ala Leu Thr Glu Ile Phe Gly Asp Asp Ser Val Leu
385                 390                 395                 400
Gln Phe Gly Gly Gly Thr Leu Gly His Pro Trp Gly Asn Ala Pro Gly
                405                 410                 415
Ala Val Ala Asn Arg Val Ala Leu Glu Ala Cys Val Gln Ala Arg Asn
            420                 425                 430
Glu Gly Arg Asp Leu Ala Arg Glu Gly Asn Glu Val Ile Arg Glu Ala
        435                 440                 445
Ala Lys Trp Ser Pro Glu Leu Ala Ala Ala Cys Glu Val Trp Lys Glu
    450                 455                 460
Ile Lys Phe Glu Phe Glu Thr Ile Asp Thr Val
465                 470                 475
```

<210> 6
<211> 475
<212> PRT
<213> Artificial

<220>
<223> 50% da sequência de consenso para charophyta (4 espécies)

<400> 6

```
Met Ser Pro Gln Thr Glu Thr Lys Ala Gly Ala Gly Phe Lys Ala Gly
1               5                   10                  15
Val Lys Asp Tyr Arg Leu Thr Tyr Tyr Thr Pro Asp Tyr Glu Thr Lys
            20                  25                  30
Glu Thr Asp Ile Leu Ala Ala Phe Arg Met Thr Pro Gln Ala Gly Val
        35                  40                  45
Pro Pro Glu Glu Ala Gly Ala Ala Val Ala Ala Glu Ser Ser Thr Gly
    50                  55                  60
Thr Trp Thr Thr Val Trp Thr Asp Gly Leu Thr Ser Leu Asp Arg Tyr
65                  70                  75                  80
Lys Gly Arg Cys Tyr Asp Ile Glu Pro Val Ala Gly Glu Glu Asn Gln
                85                  90                  95
Tyr Ile Ala Tyr Val Ala Tyr Pro Leu Asp Leu Phe Glu Glu Gly Ser
            100                 105                 110
Val Thr Asn Leu Phe Thr Ser Ile Val Gly Asn Val Phe Gly Phe Lys
        115                 120                 125
Ala Leu Arg Ala Leu Arg Leu Glu Asp Leu Arg Ile Pro Pro Ala Tyr
    130                 135                 140
Ser Lys Thr Phe Gln Gly Pro Pro His Gly Ile Gln Val Glu Arg Asp
145                 150                 155                 160
Lys Ile Asn Lys Tyr Gly Arg Pro Leu Leu Gly Cys Thr Ile Lys Pro
                165                 170                 175
Lys Leu Gly Leu Ser Ala Lys Asn Tyr Gly Arg Ala Val Tyr Glu Cys
            180                 185                 190
Leu Arg Gly Gly Leu Asp Phe Thr Lys Asp Asp Glu Asn Val Asn Ser
        195                 200                 205
Gln Pro Phe Met Arg Trp Arg Asp Arg Phe Leu Phe Val Ala Glu Ala
```

```
    210                 215                 220
Ile Phe Lys Ala Gln Ala Glu Thr Gly Glu Ile Lys Gly His Tyr Leu
225                 230                 235                 240
Asn Ala Thr Ala Gly Thr Cys Glu Glu Met Leu Lys Arg Ala Ala Tyr
                245                 250                 255
Ala Lys Glu Leu Gly Val Pro Ile Ile Met His Asp Tyr Leu Thr Gly
            260                 265                 270
Gly Phe Thr Ala Asn Thr Ser Leu Ala His Tyr Cys Arg Asp Asn Gly
        275                 280                 285
Leu Leu Leu His Ile His Arg Ala Met His Ala Val Ile Asp Arg Gln
    290                 295                 300
Lys Asn His Gly Ile His Phe Arg Val Leu Ala Lys Ala Leu Arg Met
305                 310                 315                 320
Ser Gly Gly Asp His Ile His Ser Gly Thr Val Val Gly Lys Leu Glu
                325                 330                 335
Gly Glu Arg Glu Val Thr Leu Gly Phe Val Asp Leu Leu Arg Asp Asp
            340                 345                 350
Tyr Ile Glu Lys Asp Arg Ser Arg Gly Ile Tyr Phe Thr Gln Asp Trp
        355                 360                 365
Val Ser Met Pro Gly Val Leu Pro Val Ala Ser Gly Gly Ile His Val
    370                 375                 380
Trp His Met Pro Ala Leu Thr Glu Ile Phe Gly Asp Asp Ser Val Leu
385                 390                 395                 400
Gln Phe Gly Gly Gly Thr Leu Gly His Pro Trp Gly Asn Ala Pro Gly
                405                 410                 415
Ala Val Ala Asn Arg Val Ala Leu Glu Ala Cys Val Gln Ala Arg Asn
            420                 425                 430
Glu Gly Arg Asp Leu Ala Arg Glu Gly Asn Asp Val Ile Arg Glu Ala
        435                 440                 445
Cys Lys Trp Ser Pro Glu Leu Ala Ala Ala Cys Glu Val Trp Lys Glu
    450                 455                 460
Ile Lys Phe Glu Phe Asp Thr Ile Asp Thr Leu
465                 470                 475
```

<210> 7
<211> 475
<212> PRT
<213> Artificial

<220>
<223> 50% da sequência de consenso para chlorophyta (4 espécies)

<220>
<221> MISC_FEATURE
<222> (282)..(282)
<223> Residue is one of A, C, D, G, N, P, S, T or V

<400> 7

```
Met Val Pro Gln Thr Glu Thr Lys Ala Gly Ala Gly Phe Lys Ala Gly
1               5                   10                  15
Val Lys Asp Tyr Arg Leu Thr Tyr Tyr Thr Pro Asp Tyr Val Val Lys
            20                  25                  30
Asp Thr Asp Ile Leu Ala Ala Phe Arg Met Thr Pro Gln Pro Gly Val
        35                  40                  45
Pro Pro Glu Glu Cys Gly Ala Ala Val Ala Ala Glu Ser Ser Thr Gly
    50                  55                  60
Thr Trp Thr Thr Val Trp Thr Asp Gly Leu Thr Ser Leu Asp Arg Tyr
65                  70                  75                  80
Lys Gly Arg Cys Tyr Asp Ile Glu Pro Val Pro Gly Glu Asp Asn Gln
```

```
                85                  90                  95
Tyr Ile Ala Tyr Val Ala Tyr Pro Ile Asp Leu Phe Glu Glu Gly Ser
            100                 105                 110
Val Thr Asn Leu Phe Thr Ser Ile Val Gly Asn Val Phe Gly Phe Lys
        115                 120                 125
Ala Leu Arg Ala Leu Arg Leu Glu Asp Leu Arg Ile Pro Pro Ala Tyr
    130                 135                 140
Val Lys Thr Phe Gln Gly Pro Pro His Gly Ile Gln Val Glu Arg Asp
145                 150                 155                 160
Lys Leu Asn Lys Tyr Gly Arg Gly Leu Leu Gly Cys Thr Ile Lys Pro
                165                 170                 175
Lys Leu Gly Leu Ser Ala Lys Asn Tyr Gly Arg Ala Val Tyr Glu Cys
            180                 185                 190
Leu Arg Gly Gly Leu Asp Phe Thr Lys Asp Asp Glu Asn Val Asn Ser
        195                 200                 205
Gln Pro Phe Met Arg Trp Arg Asp Arg Phe Leu Phe Val Ala Glu Ala
    210                 215                 220
Ile Tyr Lys Ala Gln Ala Glu Thr Gly Glu Val Lys Gly His Tyr Leu
225                 230                 235                 240
Asn Ala Thr Ala Ala Thr Cys Glu Glu Met Leu Lys Arg Ala Val Cys
                245                 250                 255
Ala Lys Glu Leu Gly Val Pro Ile Ile Met His Asp Tyr Leu Thr Gly
            260                 265                 270
Gly Phe Thr Ala Asn Thr Ser Leu Ala Xaa Tyr Cys Arg Asp Asn Gly
        275                 280                 285
Leu Leu Leu His Ile His Arg Ala Met His Ala Val Ile Asp Arg Gln
    290                 295                 300
Arg Asn His Gly Ile His Phe Arg Val Leu Ala Lys Ala Leu Arg Leu
305                 310                 315                 320
Ser Gly Gly Asp His Leu His Ser Gly Thr Val Val Gly Lys Leu Glu
                325                 330                 335
Gly Glu Arg Glu Val Thr Leu Gly Phe Val Asp Leu Met Arg Asp Asp
            340                 345                 350
Tyr Ile Glu Lys Asp Arg Ser Arg Gly Ile Tyr Phe Thr Gln Asp Trp
        355                 360                 365

Cys Ser Met Pro Gly Val Met Pro Val Ala Ser Gly Gly Ile His Val
    370                 375                 380
Trp His Met Pro Ala Leu Val Glu Ile Phe Gly Asp Asp Ala Cys Leu
385                 390                 395                 400
Gln Phe Gly Gly Gly Thr Leu Gly His Pro Trp Gly Asn Ala Pro Gly
                405                 410                 415
Ala Ala Ala Asn Arg Val Ala Leu Glu Ala Cys Thr Gln Ala Arg Asn
            420                 425                 430
Glu Gly Arg Asp Leu Ala Arg Glu Gly Gly Asp Val Ile Arg Ser Ala
        435                 440                 445
Cys Lys Trp Ser Pro Glu Leu Ala Ala Ala Cys Glu Val Trp Lys Glu
    450                 455                 460
Ile Lys Phe Glu Phe Asp Thr Ile Asp Lys Leu
465                 470                 475

<210>  8
<211>  475
<212>  PRT
<213>  Pinus thunbergii

<400>  8

Met Ser Pro Lys Thr Glu Thr Lys Ala Ser Val Gly Phe Lys Ala Gly
1               5                   10                  15
Val Lys Asp Tyr Arg Leu Thr Tyr Tyr Thr Pro Glu Tyr Gln Thr Lys
```

```
            20                  25                  30
Asp Thr Asp Ile Leu Ala Ala Phe Arg Val Thr Pro Gln Pro Gly Val
        35                  40                  45
Pro Pro Glu Glu Ala Gly Ala Ala Val Ala Ala Glu Ser Ser Thr Gly
    50                  55                  60
Thr Trp Thr Thr Val Trp Thr Asp Gly Leu Thr Ser Leu Asp Arg Tyr
65                  70                  75                  80
Lys Gly Arg Cys Tyr Asp Ile Glu Pro Val Pro Gly Glu Glu Thr Gln
                85                  90                  95
Phe Ile Ala Tyr Val Ala Tyr Pro Leu Asp Leu Phe Glu Glu Gly Ser
            100                 105                 110
Val Thr Asn Leu Phe Thr Ser Ile Val Gly Asn Val Phe Gly Phe Lys
        115                 120                 125
Ala Leu Arg Ala Leu Arg Leu Glu Asp Leu Arg Ile Pro Pro Ser Tyr
    130                 135                 140
Ser Lys Thr Phe Gln Gly Pro Pro His Gly Ile Gln Val Glu Arg Asp
145                 150                 155                 160
Lys Leu Asn Lys Tyr Gly Arg Pro Leu Leu Gly Cys Thr Ile Lys Pro
                165                 170                 175
Lys Leu Gly Leu Ser Ala Lys Asn Tyr Gly Arg Ala Val Tyr Glu Cys
            180                 185                 190
Leu Arg Gly Gly Leu Asp Phe Thr Lys Asp Asp Glu Asn Val Asn Ser
        195                 200                 205
Gln Pro Phe Met Arg Trp Arg Asp Arg Phe Val Phe Cys Ala Glu Ala
    210                 215                 220
Leu Asn Lys Ala Gln Ala Glu Thr Gly Glu Ile Lys Gly His Tyr Leu
225                 230                 235                 240
Asn Ala Thr Ala Gly Thr Cys Glu Glu Met Met Lys Arg Ala Ile Phe
                245                 250                 255
Ala Arg Glu Leu Gly Val Pro Ile Val Met His Asp Tyr Leu Thr Gly
            260                 265                 270
Gly Phe Thr Ala Asn Thr Ser Leu Ala His Tyr Cys Arg Asp Asn Gly
        275                 280                 285
Leu Leu Leu His Ile His Arg Ala Met His Ala Val Ile Asp Arg Gln
    290                 295                 300
Arg Asn His Gly Met His Phe Arg Val Leu Ala Lys Ala Leu Arg Met
305                 310                 315                 320
Ser Gly Gly Asp His Ile His Ala Gly Thr Val Val Gly Lys Leu Glu
                325                 330                 335
Gly Glu Arg Asp Val Thr Leu Gly Phe Val Asp Leu Leu Arg Asp Asp
            340                 345                 350
Phe Ile Glu Lys Asp Arg Ser Arg Gly Ile Tyr Phe Thr Gln Asp Trp
        355                 360                 365
Val Ser Met Pro Gly Val Leu Pro Val Ala Ser Gly Gly Ile His Val
    370                 375                 380
Trp His Met Pro Ala Leu Thr Glu Ile Phe Gly Asp Asp Ser Val Leu
385                 390                 395                 400
Gln Phe Gly Gly Gly Thr Leu Gly His Pro Trp Gly Asn Ala Pro Gly
                405                 410                 415
Ala Val Ala Asn Arg Val Ala Leu Glu Ala Cys Val Gln Ala Arg Asn
            420                 425                 430
Glu Gly Arg Asp Leu Ala Arg Glu Gly Asn Glu Val Ile Arg Glu Ala
        435                 440                 445
Cys Lys Trp Ser Pro Glu Leu Ala Ala Ala Cys Glu Ile Trp Lys Glu
    450                 455                 460
Ile Lys Phe Glu Phe Asp Val Ile Asp Arg Leu
465                 470                 475

<210>  9
<211>  475
<212>  PRT
```

<213> Equisetum arvense

<400> 9

```
Met Ser Pro Gln Thr Glu Thr Lys Ala Gly Val Gly Phe Lys Ala Gly
1               5                  10                  15
Val Lys Asp Tyr Arg Leu Thr Tyr Phe Thr Pro Asp Tyr Glu Thr Lys
            20                  25                  30
Asp Thr Asp Ile Leu Ala Ala Phe Arg Met Thr Pro Gln Pro Gly Val
        35                  40                  45
Pro Pro Glu Glu Ala Gly Ala Ala Val Ala Ala Glu Ser Ser Thr Gly
    50                  55                  60
Thr Trp Thr Thr Val Trp Thr Asp Gly Leu Thr Ser Leu Asp Arg Tyr
65                  70                  75                  80
Lys Gly Arg Cys Tyr Asn Ile Glu Pro Val Ala Gly Glu Asp Asn Gln
                85                  90                  95
Phe Ile Ala Tyr Val Ala Tyr Pro Leu Asp Leu Phe Glu Glu Gly Ser
            100                 105                 110
Val Thr Asn Leu Phe Thr Ser Ile Val Gly Asn Val Phe Gly Phe Lys
        115                 120                 125
Ala Leu Arg Ala Leu Arg Leu Glu Asp Leu Arg Ile Pro Pro Ala Tyr
    130                 135                 140
Ser Lys Thr Phe Ile Gly Pro Pro His Gly Ile Gln Val Glu Arg Asp
145                 150                 155                 160
Lys Leu Asn Lys Tyr Gly Arg Pro Leu Leu Gly Cys Thr Ile Lys Pro
                165                 170                 175
Lys Leu Gly Leu Ser Ala Lys Asn Tyr Gly Arg Ala Val Tyr Glu Cys
            180                 185                 190
Leu Arg Gly Gly Leu Asp Phe Thr Lys Asp Asp Glu Asn Val Asn Ser
        195                 200                 205
Gln Pro Phe Met Arg Trp Arg Asp Arg Phe Leu Phe Val Ala Glu Ala
    210                 215                 220
Leu Phe Lys Ser Gln Ala Glu Thr Gly Glu Ile Lys Gly His Tyr Leu
225                 230                 235                 240
Asn Ala Thr Ala Gly Thr Cys Glu Glu Met Leu Lys Arg Ala Val Phe
                245                 250                 255
Ala Arg Glu Leu Gly Ala Pro Ile Val Met His Asp Tyr Leu Thr Gly
            260                 265                 270
Gly Phe Thr Ala Asn Thr Ser Leu Ala Phe Tyr Cys Arg Asp Asn Gly
        275                 280                 285
Leu Leu Leu His Ile His Arg Ala Met His Ala Val Ile Asp Arg Gln
    290                 295                 300
Lys Asn His Gly Ile His Phe Arg Val Leu Ala Lys Ala Leu Arg Met
305                 310                 315                 320
Ser Gly Gly Asp His Ile His Thr Gly Thr Val Val Gly Lys Leu Glu
                325                 330                 335
Gly Glu Arg Asp Leu Thr Leu Gly Phe Val Asp Leu Leu Arg Asp Asp
            340                 345                 350
Phe Ile Glu Lys Asp Arg Ser Arg Gly Ile Tyr Phe Thr Gln Asp Trp
        355                 360                 365
Val Ser Met Pro Gly Val Leu Pro Val Ala Ser Gly Gly Ile His Val
    370                 375                 380
Trp His Met Pro Ala Leu Thr Glu Ile Phe Gly Asp Asp Ser Val Leu
385                 390                 395                 400
Gln Phe Gly Gly Gly Thr Leu Gly His Pro Trp Gly Asn Ala Pro Gly
                405                 410                 415
Ala Val Ala Asn Arg Val Ala Val Glu Ala Cys Val Gln Ala Arg Asn
            420                 425                 430
Glu Gly Arg Asp Leu Ala Thr Glu Gly Asn Asp Ile Ile Arg Glu Ala
        435                 440                 445
Ala Lys Trp Ser Pro Glu Leu Ala Ala Ala Cys Glu Val Trp Lys Glu
```

```
    450                 455                 460
Ile Lys Phe Glu Tyr Glu Ala Met Asp Thr Leu
465                 470                 475


<210>  10
<211>  475
<212>  PRT
<213>  Welwitschia mirabilis

<400>  10

Met Ser Pro Lys Thr Glu Thr Lys Ala Ser Val Gly Phe Gln Ala Gly
1               5                  10                  15
Val Lys Asp Tyr Arg Leu Thr Tyr Tyr Thr Pro Glu Tyr Gln Thr Lys
            20                  25                  30
Asp Thr Asp Ile Leu Ala Ala Phe Arg Val Thr Pro Gln Pro Gly Val
        35                  40                  45
Pro Pro Glu Glu Ala Gly Ala Ala Val Ala Ala Glu Ser Ser Thr Gly
    50                  55                  60
Thr Trp Thr Thr Val Trp Thr Asp Gly Leu Thr Ser Leu Asp Arg Tyr
65                  70                  75                  80
Lys Gly Arg Cys Tyr Asp Leu Glu Pro Val Pro Gly Glu Asp Asn Gln
                85                  90                  95
Tyr Ile Ala Tyr Val Ala Tyr Pro Leu Asp Leu Phe Glu Glu Gly Ser
            100                 105                 110
Val Thr Asn Met Phe Thr Ser Ile Val Gly Asn Val Phe Gly Phe Lys
        115                 120                 125
Ala Leu Arg Ala Leu Arg Leu Glu Asp Leu Arg Ile Pro Thr Ser Tyr
    130                 135                 140
Ile Lys Thr Phe Gln Gly Pro Pro His Gly Ile Gln Val Glu Arg Asp
145                 150                 155                 160
Lys Leu Asn Lys Tyr Gly Arg Pro Leu Leu Gly Cys Thr Ile Lys Pro
                165                 170                 175
Lys Leu Gly Leu Ser Ala Lys Asn Tyr Gly Arg Ala Val Tyr Glu Cys
            180                 185                 190
Leu Arg Gly Gly Leu Asp Phe Thr Lys Asp Asp Glu Asn Val Asn Ser
        195                 200                 205
Gln Pro Phe Met Arg Trp Arg Asp Arg Phe Val Phe Cys Ala Glu Ala
    210                 215                 220
Ile Tyr Lys Ala Gln Ala Glu Thr Gly Glu Ile Lys Gly His Tyr Leu
225                 230                 235                 240
Asn Ala Thr Ala Gly Thr Cys Glu Glu Met Ile Lys Arg Ala Val Phe
                245                 250                 255
Ala Arg Glu Leu Gly Val Pro Ile Val Met His Asp Tyr Leu Thr Gly
            260                 265                 270
Gly Phe Thr Ala Asn Thr Thr Leu Ala His Tyr Cys Arg Asp Asn Gly
        275                 280                 285
Leu Leu Leu His Ile His Arg Ala Met His Ala Val Ile Asp Arg Gln
    290                 295                 300
Lys Asn His Gly Met His Phe Arg Val Leu Ala Lys Ala Leu Arg Met
305                 310                 315                 320
Ser Gly Gly Asp His Ile His Ala Gly Thr Val Val Gly Lys Leu Glu
                325                 330                 335
Gly Glu Arg Glu Ile Thr Leu Gly Phe Val Asp Leu Leu Arg Asp Asp
            340                 345                 350
Phe Ile Glu Lys Asp Arg Ser Arg Gly Ile Tyr Phe Thr Gln Asp Trp
        355                 360                 365
Val Ser Met Pro Gly Val Met Pro Val Ala Ser Gly Gly Ile His Val
    370                 375                 380
Trp His Met Pro Ala Leu Thr Asp Ile Phe Gly Asp Asp Ala Val Leu
```

```
385                 390                 395                 400
Gln Phe Gly Gly Gly Thr Leu Gly His Pro Trp Gly Asn Ala Pro Gly
                405                 410                 415
Ala Val Ala Asn Arg Val Ala Leu Glu Ala Cys Val Gln Ala Arg Asn
            420                 425                 430
Glu Gly Arg Asp Leu Ala Arg Glu Gly Asn Glu Val Ile Arg Glu Ala
        435                 440                 445
Ala Lys Trp Ser Pro Glu Leu Ala Ala Ala Cys Glu Val Trp Lys Glu
    450                 455                 460
Ile Lys Phe Glu Phe Glu Ser Val Asp Thr Leu
465                 470                 475
```

<210> 11
<211> 475
<212> PRT
<213> Artificial

<220>
<223> 50% da sequência de consenso para magnoliophyta (134 espécies)

<220>
<221> MISC_FEATURE
<222> (142)..(142)
<223> Residue is one of A, C, D, G, N, P, S, T or V

<220>
<221> MISC_FEATURE
<222> (251)..(251)
<223> Residue is one of I, L or V

<400> 11

```
Met Ser Pro Gln Thr Glu Thr Lys Ala Ser Val Gly Phe Lys Ala Gly
1               5                   10                  15
Val Lys Asp Tyr Lys Leu Thr Tyr Tyr Thr Pro Asp Tyr Glu Thr Lys
            20                  25                  30
Asp Thr Asp Ile Leu Ala Ala Phe Arg Val Thr Pro Gln Pro Gly Val
        35                  40                  45
Pro Pro Glu Glu Ala Gly Ala Ala Val Ala Ala Glu Ser Ser Thr Gly
    50                  55                  60
Thr Trp Thr Thr Val Trp Thr Asp Gly Leu Thr Ser Leu Asp Arg Tyr
65                  70                  75                  80
Lys Gly Arg Cys Tyr His Ile Glu Pro Val Ala Gly Glu Glu Asn Gln
                85                  90                  95
Phe Ile Ala Tyr Val Ala Tyr Pro Leu Asp Leu Phe Glu Glu Gly Ser
            100                 105                 110
Val Thr Asn Met Phe Thr Ser Ile Val Gly Asn Val Phe Gly Phe Lys
        115                 120                 125
Ala Leu Arg Ala Leu Arg Leu Glu Asp Leu Arg Ile Pro Xaa Ala Tyr
    130                 135                 140
Val Lys Thr Phe Gln Gly Pro Pro His Gly Ile Gln Val Glu Arg Asp
145                 150                 155                 160
Lys Leu Asn Lys Tyr Gly Arg Pro Leu Leu Gly Cys Thr Ile Lys Pro
                165                 170                 175
Lys Leu Gly Leu Ser Ala Lys Asn Tyr Gly Arg Ala Val Tyr Glu Cys
            180                 185                 190
Leu Arg Gly Gly Leu Asp Phe Thr Lys Asp Asp Glu Asn Val Asn Ser
        195                 200                 205
Gln Pro Phe Met Arg Trp Arg Asp Arg Phe Leu Phe Cys Ala Glu Ala
    210                 215                 220
```

```
Ile Tyr Lys Ala Gln Ala Glu Thr Gly Glu Ile Lys Gly His Tyr Leu
225                 230                 235                 240
Asn Ala Thr Ala Gly Thr Cys Glu Glu Met Xaa Lys Arg Ala Val Phe
                245                 250                 255
Ala Arg Glu Leu Gly Val Pro Ile Val Met His Asp Tyr Leu Thr Gly
            260                 265                 270
Gly Phe Thr Ala Asn Thr Ser Leu Ala His Tyr Cys Arg Asp Asn Gly
        275                 280                 285
Leu Leu Leu His Ile His Arg Ala Met His Ala Val Ile Asp Arg Gln
    290                 295                 300
Lys Asn His Gly Met His Phe Arg Val Leu Ala Lys Ala Leu Arg Met
305                 310                 315                 320
Ser Gly Gly Asp His Ile His Ala Gly Thr Val Val Gly Lys Leu Glu
                325                 330                 335
Gly Glu Arg Glu Ile Thr Leu Gly Phe Val Asp Leu Leu Arg Asp Asp
            340                 345                 350
Phe Ile Glu Lys Asp Arg Ser Arg Gly Ile Tyr Phe Thr Gln Asp Trp
        355                 360                 365
Val Ser Leu Pro Gly Val Leu Pro Val Ala Ser Gly Gly Ile His Val
    370                 375                 380
Trp His Met Pro Ala Leu Thr Glu Ile Phe Gly Asp Asp Ser Val Leu
385                 390                 395                 400
Gln Phe Gly Gly Gly Thr Leu Gly His Pro Trp Gly Asn Ala Pro Gly
                405                 410                 415
Ala Val Ala Asn Arg Val Ala Leu Glu Ala Cys Val Gln Ala Arg Asn
            420                 425                 430
Glu Gly Arg Asp Leu Ala Arg Glu Gly Asn Glu Ile Ile Arg Glu Ala
        435                 440                 445
Ser Lys Trp Ser Pro Glu Leu Ala Ala Ala Cys Glu Val Trp Lys Glu
    450                 455                 460
Ile Lys Phe Glu Phe Glu Ala Met Asp Thr Leu
465                 470                 475
```

<210> 12
<211> 475
<212> PRT
<213> Artificial

<220>
<223> 50% da sequência de consenso para pinophyta (2 espécies)

<400> 12

```
Met Ser Pro Lys Thr Glu Thr Lys Ala Ser Val Gly Phe Lys Ala Gly
1               5                   10                  15
Val Lys Asp Tyr Arg Leu Thr Tyr Tyr Thr Pro Glu Tyr Gln Thr Lys
            20                  25                  30
Asp Thr Asp Ile Leu Ala Ala Phe Arg Val Thr Pro Gln Pro Gly Val
        35                  40                  45
Pro Ala Glu Glu Ala Gly Ala Ala Val Ala Ala Glu Ser Ser Thr Gly
    50                  55                  60
Thr Trp Thr Thr Val Trp Thr Asp Gly Leu Thr Ser Leu Asp Arg Tyr
65                  70                  75                  80
Lys Gly Arg Cys Tyr Asp Ile Glu Pro Val Pro Gly Glu Glu Asn Gln
                85                  90                  95
Phe Ile Ala Tyr Val Ala Tyr Pro Leu Asp Leu Phe Glu Glu Gly Ser
            100                 105                 110
Val Thr Asn Leu Phe Thr Ser Ile Val Gly Asn Val Phe Gly Phe Lys
        115                 120                 125
Ala Leu Arg Ala Leu Arg Leu Glu Asp Leu Arg Ile Pro Pro Ala Tyr
```

```
    130                 135                 140
Ser Lys Thr Phe Gln Gly Pro Pro His Gly Ile Gln Val Glu Arg Asp
145                 150                 155                 160
Lys Leu Asn Lys Tyr Gly Arg Pro Leu Leu Gly Cys Thr Ile Lys Pro
                165                 170                 175
Lys Leu Gly Leu Ser Ala Lys Asn Tyr Gly Arg Ala Val Tyr Glu Cys
            180                 185                 190
Leu Arg Gly Gly Leu Asp Phe Thr Lys Asp Asp Glu Asn Val Asn Ser
        195                 200                 205
Gln Pro Phe Met Arg Trp Arg Asp Arg Phe Val Phe Cys Ala Glu Ala
    210                 215                 220
Ile Asn Lys Ala Gln Ala Glu Thr Gly Glu Ile Lys Gly His Tyr Leu
225                 230                 235                 240
Asn Ala Thr Ala Gly Thr Cys Glu Glu Met Met Lys Arg Ala Ile Phe
                245                 250                 255
Ala Arg Glu Leu Gly Val Pro Ile Val Met His Asp Tyr Leu Thr Gly
            260                 265                 270
Gly Phe Thr Ala Asn Thr Ser Leu Ala His Tyr Cys Arg Asp Asn Gly
        275                 280                 285
Leu Leu Leu His Ile His Arg Ala Met His Ala Val Ile Asp Arg Gln
    290                 295                 300
Lys Asn His Gly Met His Phe Arg Val Leu Ala Lys Ala Leu Arg Met
305                 310                 315                 320
Ser Gly Gly Asp His Ile His Ala Gly Thr Val Val Gly Lys Leu Glu
                325                 330                 335
Gly Glu Arg Asp Val Thr Leu Gly Phe Val Asp Leu Leu Arg Asp Asp
            340                 345                 350
Phe Ile Glu Lys Asp Arg Ser Arg Gly Ile Tyr Phe Thr Gln Asp Trp
        355                 360                 365
Val Ser Met Pro Gly Val Leu Pro Val Ala Ser Gly Gly Ile His Val
    370                 375                 380
Trp His Met Pro Ala Leu Thr Glu Ile Phe Gly Asp Asp Ser Val Leu
385                 390                 395                 400
Gln Phe Gly Gly Gly Thr Leu Gly His Pro Trp Gly Asn Ala Pro Gly
                405                 410                 415
Ala Val Ala Asn Arg Val Ala Leu Glu Ala Cys Val Gln Ala Arg Asn
            420                 425                 430
Glu Gly Arg Asp Leu Ala Arg Glu Gly Asn Glu Val Ile Arg Glu Ala
        435                 440                 445
Cys Lys Trp Ser Pro Glu Leu Ala Ala Ala Cys Glu Ile Trp Lys Glu
    450                 455                 460
Ile Lys Phe Glu Phe Asp Ala Ile Asp Arg Leu
465                 470                 475
```

<210> 13
<211> 475
<212> PRT
<213> Artificial

<220>
<223> 50% da sequência de consenso para pteridophyta (2 espécies)

<400> 13

```
Met Ser Pro Gln Thr Glu Thr Lys Thr Gly Ile Gly Phe Lys Ala Gly
1               5                   10                  15
Val Lys Asp Tyr Arg Leu Thr Tyr Tyr Thr Pro Asp Tyr Glu Thr Lys
            20                  25                  30
Asp Thr Asp Ile Leu Ala Ala Phe Arg Met Thr Pro Gln Pro Gly Val
        35                  40                  45
```

```
Pro Ala Glu Glu Ala Gly Ala Ala Val Ala Ala Glu Ser Ser Thr Gly
    50                  55                  60
Thr Trp Thr Thr Val Trp Thr Asp Gly Leu Thr Ser Leu Asp Arg Tyr
65                  70                  75                  80
Lys Gly Arg Cys Tyr Asp Ile Glu Pro Val Ala Gly Glu Glu Asn Gln
                85                  90                  95
Tyr Ile Ala Tyr Val Ala Tyr Pro Leu Asp Leu Phe Glu Glu Gly Ser
            100                 105                 110
Val Thr Asn Met Leu Thr Ser Ile Val Gly Asn Val Phe Gly Phe Lys
        115                 120                 125
Ala Leu Arg Ala Leu Arg Leu Glu Asp Leu Arg Ile Leu Pro Ala Tyr
    130                 135                 140
Ser Lys Thr Phe Ile Gly Pro Pro His Gly Ile Gln Val Glu Arg Asp
145                 150                 155                 160
Lys Leu Asn Lys Tyr Gly Arg Pro Leu Leu Gly Cys Thr Ile Lys Pro
                165                 170                 175
Lys Leu Gly Leu Cys Ala Lys Asn Tyr Gly Arg Ala Val Tyr Glu Cys
            180                 185                 190
Leu Arg Gly Gly Leu Asp Phe Thr Lys Asp Asp Glu Asn Val Asn Ser
        195                 200                 205
Gln Pro Phe Met Arg Trp Arg Asp Arg Phe Leu Phe Val Ala Glu Ala
    210                 215                 220
Leu Phe Lys Ala Gln Ala Glu Thr Gly Glu Ile Lys Gly His Tyr Leu
225                 230                 235                 240
Asn Ala Thr Ala Ala Thr Cys Glu Glu Met Ile Lys Arg Ala Ile Phe
                245                 250                 255
Ala Arg Glu Leu Gly Ala Pro Ile Val Met His Asp Tyr Leu Thr Gly
            260                 265                 270
Gly Phe Thr Ala Asn Thr Ser Leu Ala Phe Tyr Cys Arg Asp Asn Gly
        275                 280                 285
Leu Leu Leu His Ile His Arg Ala Met His Ala Val Ile Asp Arg Gln
    290                 295                 300
Lys Asn His Gly Met His Phe Arg Val Leu Ala Lys Ala Leu Arg Met
305                 310                 315                 320
Ser Gly Gly Asp His Ile His Ala Gly Thr Val Val Gly Lys Leu Glu
                325                 330                 335
Gly Glu Arg Glu Val Thr Leu Gly Phe Val Asp Leu Leu Arg Asp Asp
            340                 345                 350
Tyr Ile Glu Lys Asp Arg Ser Arg Gly Ile Tyr Phe Thr Gln Asp Trp
        355                 360                 365
Val Ser Met Pro Gly Val Phe Pro Val Ala Ser Gly Gly Ile His Val
    370                 375                 380
Trp His Met Pro Ala Leu Thr Glu Ile Phe Gly Asp Asp Ser Val Leu
385                 390                 395                 400
Gln Phe Gly Gly Gly Thr Leu Gly His Pro Trp Gly Asn Ala Pro Gly
                405                 410                 415
Ala Val Ala Asn Arg Val Ala Leu Glu Ala Cys Val Gln Ala Arg Asn
            420                 425                 430
Glu Gly Arg Asp Leu Ala Arg Glu Gly Asn Glu Ile Ile Arg Glu Ala
        435                 440                 445
Ser Lys Trp Ser Pro Glu Leu Ala Ala Ala Cys Glu Val Trp Lys Glu
    450                 455                 460
Ile Lys Phe Glu Phe Glu Thr Ile Asp Thr Val
465                 470                 475

<210>  14
<211>  493
<212>  PRT
<213>  Galdieria partita

<400>  14
```

```
Met Ser Gln Ser Ile Glu Glu Lys Ser Val Gln Glu Arg Thr Arg Ile
1               5                  10                  15
Lys Asn Ser Arg Tyr Glu Ser Gly Val Ile Pro Tyr Ala Lys Met Gly
            20                  25                  30
Tyr Trp Asn Pro Asp Tyr Gln Val Lys Asp Thr Asp Val Leu Ala Leu
        35                  40                  45
Phe Arg Val Thr Pro Gln Pro Gly Val Asp Pro Ile Glu Ala Ala Ala
    50                  55                  60
Ala Val Ala Gly Glu Ser Ser Thr Ala Thr Trp Thr Val Val Trp Thr
65                  70                  75                  80
Asp Leu Leu Thr Ala Ala Asp Leu Tyr Arg Ala Lys Ala Tyr Lys Val
                85                  90                  95
Asp Gln Val Pro Asn Asn Pro Glu Gln Tyr Phe Ala Tyr Ile Ala Tyr
            100                 105                 110
Glu Leu Asp Leu Phe Glu Glu Gly Ser Ile Ala Asn Leu Thr Ala Ser
        115                 120                 125
Ile Ile Gly Asn Val Phe Gly Phe Lys Ala Val Lys Ala Leu Arg Leu
    130                 135                 140
Glu Asp Met Arg Leu Pro Leu Ala Tyr Leu Lys Thr Phe Gln Gly Pro
145                 150                 155                 160
Ala Thr Gly Val Ile Leu Glu Arg Glu Arg Leu Asp Lys Phe Gly Arg
                165                 170                 175
Pro Leu Leu Gly Cys Thr Thr Lys Pro Lys Leu Gly Leu Ser Gly Lys
            180                 185                 190
Asn Tyr Gly Arg Val Val Tyr Glu Ala Leu Lys Gly Gly Leu Asp Phe
        195                 200                 205
Val Lys Asp Asp Glu Asn Ile Asn Ser Gln Pro Phe Met Arg Trp Arg
    210                 215                 220
Glu Arg Tyr Leu Phe Thr Met Glu Ala Val Asn Lys Ala Ser Ala Ala
225                 230                 235                 240
Thr Gly Glu Val Lys Gly His Tyr Leu Asn Val Thr Ala Ala Thr Met
                245                 250                 255
Glu Glu Met Tyr Ala Arg Ala Asn Phe Ala Lys Glu Leu Gly Ser Val
            260                 265                 270
Ile Ile Met Ile Asp Leu Val Ile Gly Tyr Thr Ala Ile Gln Thr Met
        275                 280                 285
Ala Lys Trp Ala Arg Asp Asn Asp Met Ile Leu His Leu His Arg Ala
    290                 295                 300
Gly Asn Ser Thr Tyr Ser Arg Gln Lys Asn His Gly Met Asn Phe Arg
305                 310                 315                 320
Val Ile Cys Lys Trp Met Arg Met Ala Gly Val Asp His Ile His Ala
                325                 330                 335
Gly Thr Val Val Gly Lys Leu Glu Gly Asp Pro Ile Ile Thr Arg Gly
            340                 345                 350
Phe Tyr Lys Thr Leu Leu Leu Pro Lys Leu Glu Arg Asn Leu Gln Glu
        355                 360                 365
Gly Leu Phe Phe Asp Met Glu Trp Ala Ser Leu Arg Lys Val Met Pro
    370                 375                 380
Val Ala Ser Gly Gly Ile His Ala Gly Gln Met His Gln Leu Ile His
385                 390                 395                 400
Tyr Leu Gly Glu Asp Val Val Leu Gln Phe Gly Gly Gly Thr Ile Gly
                405                 410                 415
His Pro Asp Gly Ile Gln Ala Gly Ala Thr Ala Asn Arg Val Ala Leu
            420                 425                 430
Glu Ala Met Ile Leu Ala Arg Asn Glu Asn Arg Asp Tyr Leu Thr Glu
        435                 440                 445
Gly Pro Glu Ile Leu Arg Glu Ala Ala Lys Thr Cys Gly Ala Leu Arg
    450                 455                 460
Thr Ala Leu Asp Leu Trp Lys Asp Ile Thr Phe Asn Tyr Thr Ser Thr
```

```
465                 470                 475                 480
Asp Thr Ser Asp Phe Val Glu Thr Pro Thr Ala Asn Ile
                485                 490
```

<210> 15
<211> 488
<212> PRT
<213> Griffithsia monolis

<400> 15

```
Met Ser Asn Ser Val Glu Glu Arg Thr Arg Ile Lys Asn Glu Arg Tyr
1               5                  10                  15
Glu Ser Gly Val Ile Pro Tyr Ala Lys Met Gly Tyr Trp Asp Pro Asn
            20                  25                  30
Tyr Ala Val Lys Asp Thr Asp Ile Leu Ala Leu Phe Arg Val Ser Pro
        35                  40                  45
Gln Pro Gly Val Asp Pro Val Glu Ala Ser Ala Ala Val Ala Gly Glu
    50                  55                  60
Ser Ser Thr Ala Thr Trp Thr Val Val Trp Thr Asp Leu Leu Thr Ala
65                  70                  75                  80
Cys Asp Leu Tyr Arg Ala Lys Ala Tyr Lys Val Glu Ser Val Pro Asn
                85                  90                  95
Thr Ser Asp Gln Tyr Phe Ala Tyr Ile Ser Tyr Asp Ile Asp Leu Phe
            100                 105                 110
Glu Glu Gly Ser Ile Ala Asn Leu Thr Ala Ser Ile Ile Gly Asn Val
        115                 120                 125
Phe Gly Phe Lys Ala Val Lys Ala Leu Arg Leu Glu Asp Met Arg Ile
    130                 135                 140
Pro Val Ala Tyr Leu Lys Thr Phe Gln Gly Pro Ala Thr Gly Ile Val
145                 150                 155                 160
Val Glu Arg Glu Arg Met Asp Lys Phe Gly Arg Pro Phe Leu Gly Ala
                165                 170                 175
Thr Val Lys Pro Lys Leu Gly Leu Ser Gly Lys Asn Tyr Gly Arg Val
            180                 185                 190
Val Tyr Glu Gly Leu Arg Gly Gly Leu Asp Phe Leu Lys Asp Asp Glu
        195                 200                 205
Asn Ile Asn Ser Gln Pro Phe Met Arg Trp Lys Glu Arg Phe Leu Tyr
    210                 215                 220
Ser Ile Glu Ala Val Asn Arg Ser Ile Ala Ala Thr Gly Glu Val Lys
225                 230                 235                 240
Gly His Tyr Met Asn Val Thr Ala Ala Thr Met Glu Glu Met Tyr Glu
                245                 250                 255
Arg Ala Glu Phe Ala Lys Gln Leu Gly Thr Val Ile Ile Met Ile Asp
            260                 265                 270
Leu Val Ile Gly Tyr Thr Ala Ile Gln Thr Met Gly Ile Trp Ala Arg
        275                 280                 285
Lys Asn Asp Met Ile Leu His Leu His Arg Ala Gly Asn Ser Thr Tyr
    290                 295                 300
Ser Arg Gln Lys Ile His Gly Met Asn Phe Arg Val Ile Cys Lys Trp
305                 310                 315                 320
Met Arg Met Ala Gly Val Asp His Ile His Ala Gly Thr Val Val Gly
                325                 330                 335
Lys Leu Glu Gly Asp Pro Leu Met Ile Arg Gly Phe Tyr Asn Thr Leu
            340                 345                 350
Leu Leu Pro Tyr Leu Glu Val Asn Leu Pro Gln Gly Ile Phe Phe Gln
        355                 360                 365
Gln Asp Trp Ala Ser Leu Arg Lys Val Thr Pro Val Ala Ser Gly Gly
    370                 375                 380
Ile His Cys Gly Gln Met His Gln Leu Leu Asp Tyr Leu Gly Asn Asp
385                 390                 395                 400
```

```
Val Val Leu Gln Phe Gly Gly Gly Thr Ile Gly His Pro Asp Gly Ile
                405                 410                 415
Gln Ala Gly Ala Thr Ala Asn Arg Val Ala Leu Glu Ser Met Val Ile
            420                 425                 430
Ala Arg Asn Glu Gly Arg Asp Tyr Val Ala Glu Gly Pro Gln Ile Leu
        435                 440                 445
Arg Asp Ala Ala Lys Thr Cys Gly Pro Leu Gln Thr Ala Leu Asp Leu
    450                 455                 460
Trp Lys Asp Ile Thr Phe Asn Tyr Thr Ser Thr Asp Thr Ala Asp Phe
465                 470                 475                 480
Val Glu Thr Pro Thr Ala Asn Val
                485
```

<210> 16
<211> 472
<212> PRT
<213> Synechococcus elongatus PCC6301

<400> 16

```
Met Pro Lys Thr Gln Ser Ala Ala Gly Tyr Lys Ala Gly Val Lys Asp
1               5                   10                  15
Tyr Lys Leu Thr Tyr Tyr Thr Pro Asp Tyr Thr Pro Lys Asp Thr Asp
            20                  25                  30
Leu Leu Ala Ala Phe Arg Phe Ser Pro Gln Pro Gly Val Pro Ala Asp
        35                  40                  45
Glu Ala Gly Ala Ala Ile Ala Ala Glu Ser Ser Thr Gly Thr Trp Thr
    50                  55                  60
Thr Val Trp Thr Asp Leu Leu Thr Asp Met Asp Arg Tyr Lys Gly Lys
65                  70                  75                  80
Cys Tyr His Ile Glu Pro Val Gln Gly Glu Glu Asn Ser Tyr Phe Ala
                85                  90                  95
Phe Ile Ala Tyr Pro Leu Asp Leu Phe Glu Glu Gly Ser Val Thr Asn
            100                 105                 110
Ile Leu Thr Ser Ile Val Gly Asn Val Phe Gly Phe Lys Ala Ile Arg
        115                 120                 125
Ser Leu Arg Leu Glu Asp Ile Arg Phe Pro Val Ala Leu Val Lys Thr
    130                 135                 140
Phe Gln Gly Pro Pro His Gly Ile Gln Val Glu Arg Asp Leu Leu Asn
145                 150                 155                 160
Lys Tyr Gly Arg Pro Met Leu Gly Cys Thr Ile Lys Pro Lys Leu Gly
                165                 170                 175
Leu Ser Ala Lys Asn Tyr Gly Arg Ala Val Tyr Glu Cys Leu Arg Gly
            180                 185                 190
Gly Leu Asp Phe Thr Lys Asp Asp Glu Asn Ile Asn Ser Gln Pro Phe
        195                 200                 205
Gln Arg Trp Arg Asp Arg Phe Leu Phe Val Ala Asp Ala Ile His Lys
    210                 215                 220
Ser Gln Ala Glu Thr Gly Glu Ile Lys Gly His Tyr Leu Asn Val Thr
225                 230                 235                 240
Ala Pro Thr Cys Glu Glu Met Met Lys Arg Ala Glu Phe Ala Lys Glu
                245                 250                 255
Leu Gly Met Pro Ile Ile Met His Asp Phe Leu Thr Ala Gly Phe Thr
            260                 265                 270
Ala Asn Thr Thr Leu Ala Lys Trp Cys Arg Asp Asn Gly Val Leu Leu
        275                 280                 285
His Ile His Arg Ala Met His Ala Val Ile Asp Arg Gln Arg Asn His
    290                 295                 300
Gly Ile His Phe Arg Val Leu Ala Lys Cys Leu Arg Leu Ser Gly Gly
305                 310                 315                 320
```

```
Asp His Leu His Ser Gly Thr Val Val Gly Lys Leu Glu Gly Asp Lys
                325                 330                 335
Ala Ser Thr Leu Gly Phe Val Asp Leu Met Arg Glu Asp His Ile Glu
            340                 345                 350
Ala Asp Arg Ser Arg Gly Val Phe Phe Thr Gln Asp Trp Ala Ser Met
        355                 360                 365
Pro Gly Val Leu Pro Val Ala Ser Gly Gly Ile His Val Trp His Met
    370                 375                 380
Pro Ala Leu Val Glu Ile Phe Gly Asp Asp Ser Val Leu Gln Phe Gly
385                 390                 395                 400
Gly Gly Thr Leu Gly His Pro Trp Gly Asn Ala Pro Gly Ala Thr Ala
                405                 410                 415
Asn Arg Val Ala Leu Glu Ala Cys Val Gln Ala Arg Asn Glu Gly Arg
            420                 425                 430
Asp Leu Tyr Arg Glu Gly Gly Asp Ile Leu Arg Glu Ala Gly Lys Trp
        435                 440                 445
Ser Pro Glu Leu Ala Ala Ala Leu Asp Leu Trp Lys Glu Ile Lys Phe
    450                 455                 460
Glu Phe Glu Thr Met Asp Lys Leu
465                 470
```

<210> 17
<211> 475
<212> PRT
<213> Spinacia oleracea

<400> 17

```
Met Ser Pro Gln Thr Glu Thr Lys Ala Ser Val Gly Phe Lys Ala Gly
1               5                   10                  15
Val Lys Asp Tyr Lys Leu Thr Tyr Tyr Thr Pro Glu Tyr Glu Thr Leu
            20                  25                  30
Asp Thr Asp Ile Leu Ala Ala Phe Arg Val Ser Pro Gln Pro Gly Val
        35                  40                  45
Pro Pro Glu Glu Ala Gly Ala Ala Val Ala Ala Glu Ser Ser Thr Gly
    50                  55                  60
Thr Trp Thr Thr Val Trp Thr Asp Gly Leu Thr Asn Leu Asp Arg Tyr
65                  70                  75                  80
Lys Gly Arg Cys Tyr His Ile Glu Pro Val Ala Gly Glu Glu Asn Gln
                85                  90                  95
Tyr Ile Cys Tyr Val Ala Tyr Pro Leu Asp Leu Phe Glu Glu Gly Ser
            100                 105                 110
Val Thr Asn Met Phe Thr Ser Ile Val Gly Asn Val Phe Gly Phe Lys
        115                 120                 125
Ala Leu Arg Ala Leu Arg Leu Glu Asp Leu Arg Ile Pro Val Ala Tyr
    130                 135                 140
Val Lys Thr Phe Gln Gly Pro Pro His Gly Ile Gln Val Glu Arg Asp
145                 150                 155                 160
Lys Leu Asn Lys Tyr Gly Arg Pro Leu Leu Gly Cys Thr Ile Lys Pro
                165                 170                 175
Lys Leu Gly Leu Ser Ala Lys Asn Tyr Gly Arg Ala Val Tyr Glu Cys
            180                 185                 190
Leu Arg Gly Gly Leu Asp Phe Thr Lys Asp Asp Glu Asn Val Asn Ser
        195                 200                 205
Gln Pro Phe Met Arg Trp Arg Asp Arg Phe Leu Phe Cys Ala Glu Ala
    210                 215                 220
Leu Tyr Lys Ala Gln Ala Glu Thr Gly Glu Ile Lys Gly His Tyr Leu
225                 230                 235                 240
Asn Ala Thr Ala Gly Thr Cys Glu Asp Met Met Lys Arg Ala Val Phe
                245                 250                 255
Ala Arg Glu Leu Gly Val Pro Ile Val Met His Asp Tyr Leu Thr Gly
```

```
            260                 265                 270
Gly Phe Thr Ala Asn Thr Thr Leu Ser His Tyr Cys Arg Asp Asn Gly
        275                 280                 285
Leu Leu Leu His Ile His Arg Ala Met His Ala Val Ile Asp Arg Gln
    290                 295                 300
Lys Asn His Gly Met His Phe Arg Val Leu Ala Lys Ala Leu Arg Leu
305                 310                 315                 320

Ser Gly Gly Asp His Ile His Ser Gly Thr Val Val Gly Lys Leu Glu
                325                 330                 335
Gly Glu Arg Asp Ile Thr Leu Gly Phe Val Asp Leu Leu Arg Asp Asp
            340                 345                 350
Tyr Thr Glu Lys Asp Arg Ser Arg Gly Ile Tyr Phe Thr Gln Ser Trp
        355                 360                 365
Val Ser Thr Pro Gly Val Leu Pro Val Ala Ser Gly Gly Ile His Val
    370                 375                 380
Trp His Met Pro Ala Leu Thr Glu Ile Phe Gly Asp Asp Ser Val Leu
385                 390                 395                 400
Gln Phe Gly Gly Gly Thr Leu Gly His Pro Trp Gly Asn Ala Pro Gly
                405                 410                 415
Ala Val Ala Asn Arg Val Ala Leu Glu Ala Cys Val Gln Ala Arg Asn
            420                 425                 430
Glu Gly Arg Asp Leu Ala Arg Glu Gly Asn Thr Ile Ile Arg Glu Ala
        435                 440                 445
Thr Lys Trp Ser Pro Glu Leu Ala Ala Ala Cys Glu Val Trp Lys Glu
    450                 455                 460
Ile Lys Phe Glu Phe Pro Ala Met Asp Thr Val
465                 470                 475

<210>  18
<211>  475
<212>  PRT
<213>  Nicotiana tabacum

<400>  18

Met Ser Pro Gln Thr Glu Thr Lys Ala Ser Val Gly Phe Lys Ala Gly
1               5                   10                  15
Val Lys Asp Tyr Lys Leu Thr Tyr Tyr Thr Pro Asp Tyr Glu Thr Lys
            20                  25                  30
Asp Thr Asp Ile Leu Ala Ala Phe Arg Val Thr Pro Gln Pro Gly Val
        35                  40                  45
Pro Pro Glu Glu Ala Gly Ala Ala Val Ala Ala Glu Ser Ser Thr Gly
    50                  55                  60
Thr Trp Thr Thr Val Trp Thr Asp Gly Leu Thr Ser Leu Asp Arg Tyr
65                  70                  75                  80
Lys Gly Arg Cys Tyr His Ile Glu Pro Val Ala Gly Asp Glu Asn Gln
                85                  90                  95
Phe Ile Ala Tyr Val Ala Tyr Pro Leu Asp Leu Phe Glu Glu Gly Ser
            100                 105                 110
Val Thr Asn Met Phe Thr Ser Ile Val Gly Asn Val Phe Gly Phe Lys
        115                 120                 125
Ala Leu Arg Ala Leu Arg Leu Glu Asp Leu Arg Ile Pro Pro Ala Tyr
    130                 135                 140
Ala Lys Thr Phe Gln Gly Pro Pro His Gly Ile Gln Val Glu Arg Asp
145                 150                 155                 160
Lys Leu Asn Lys Tyr Gly Arg Pro Leu Leu Gly Cys Thr Ile Lys Pro
                165                 170                 175
Lys Leu Gly Leu Ser Ala Lys Asn Tyr Gly Arg Ala Val Tyr Glu Cys
            180                 185                 190
```

```
Leu Arg Gly Gly Leu Asp Phe Thr Lys Asp Asp Glu Asn Val Asn Ser
        195                 200                 205
Gln Pro Phe Met Arg Trp Arg Asp Arg Phe Leu Phe Cys Ala Glu Ala
    210                 215                 220
Ile Tyr Lys Ala Gln Ala Glu Thr Gly Glu Ile Lys Gly His Tyr Leu
225                 230                 235                 240
Asn Ala Thr Ala Gly Thr Cys Glu Glu Met Ile Lys Arg Ala Val Phe
                245                 250                 255
Ala Arg Glu Leu Gly Val Pro Ile Val Met His Asp Tyr Ile Thr Gly
            260                 265                 270
Gly Phe Thr Ala Asn Thr Ser Leu Ala His Tyr Cys Arg Asp Asn Gly
        275                 280                 285
Leu Leu Leu His Ile His Arg Ala Met His Ala Val Ile Asp Arg Gln
    290                 295                 300
Lys Asn His Gly Ile His Phe Arg Val Leu Ala Lys Ala Leu Arg Met
305                 310                 315                 320
Ser Gly Gly Asp His Ile His Ser Gly Thr Val Val Gly Lys Leu Glu
                325                 330                 335
Gly Glu Arg Asp Ile Thr Leu Gly Phe Val Asp Leu Leu Arg Asp Asp
            340                 345                 350
Phe Ile Glu Lys Asp Arg Ser Arg Gly Ile Tyr Phe Thr Gln Asp Trp
        355                 360                 365
Val Ser Leu Pro Gly Val Leu Pro Val Ala Ser Gly Gly Ile His Val
    370                 375                 380
Trp His Met Pro Ala Leu Thr Glu Ile Phe Gly Asp Asp Ser Val Leu
385                 390                 395                 400
Gln Phe Gly Gly Gly Thr Leu Gly His Pro Trp Gly Asn Ala Pro Gly
                405                 410                 415
Ala Val Ala Asn Arg Val Ala Leu Glu Ala Cys Val Gln Ala Arg Asn
            420                 425                 430
Glu Gly Arg Asp Leu Ala Arg Glu Gly Asn Glu Ile Ile Arg Glu Ala
        435                 440                 445
Ser Lys Trp Ser Pro Glu Leu Ala Ala Ala Cys Glu Val Trp Lys Glu
    450                 455                 460
Ile Lys Phe Glu Phe Glu Ala Met Asp Thr Leu
465                 470                 475
```

<210> 19
<211> 477
<212> PRT
<213> Oriza sativa

<400> 19

```
Met Ser Pro Gln Thr Glu Thr Lys Ala Ser Val Gly Phe Lys Ala Gly
1               5                   10                  15
Val Lys Asp Tyr Lys Leu Thr Tyr Tyr Thr Pro Glu Tyr Glu Thr Lys
            20                  25                  30
Asp Thr Asp Ile Leu Ala Ala Phe Arg Val Thr Pro Gln Pro Gly Val
        35                  40                  45
Pro Pro Glu Glu Ala Gly Ala Ala Val Ala Ala Glu Ser Ser Thr Gly
    50                  55                  60
Thr Trp Thr Thr Val Trp Thr Asp Gly Leu Thr Ser Leu Asp Arg Tyr
65                  70                  75                  80
Lys Gly Arg Cys Tyr His Ile Glu Pro Val Val Gly Glu Asp Asn Gln
                85                  90                  95
Tyr Ile Ala Tyr Val Ala Tyr Pro Leu Asp Leu Phe Glu Glu Gly Ser
            100                 105                 110
Val Thr Asn Met Phe Thr Ser Ile Val Gly Asn Val Phe Gly Phe Lys
        115                 120                 125
Ala Leu Arg Ala Leu Arg Leu Glu Asp Leu Arg Ile Pro Pro Thr Tyr
```

```
    130                 135                 140
Ser Lys Thr Phe Gln Gly Pro Pro His Gly Ile Gln Val Glu Arg Asp
145                 150                 155                 160
Lys Leu Asn Lys Tyr Gly Arg Pro Leu Leu Gly Cys Thr Ile Lys Pro
                165                 170                 175
Lys Leu Gly Leu Ser Ala Lys Asn Tyr Gly Arg Ala Cys Tyr Glu Cys
            180                 185                 190
Leu Arg Gly Gly Leu Asp Phe Thr Lys Asp Asp Glu Asn Val Asn Ser
        195                 200                 205
Gln Pro Phe Met Arg Trp Arg Asp Arg Phe Val Phe Cys Ala Glu Ala
    210                 215                 220
Ile Tyr Lys Ser Gln Ala Glu Thr Gly Glu Ile Lys Gly His Tyr Leu
225                 230                 235                 240
Asn Ala Thr Ala Gly Thr Cys Glu Glu Met Ile Lys Arg Ala Val Phe
                245                 250                 255
Ala Arg Glu Leu Gly Val Pro Ile Val Met His Asp Tyr Leu Thr Gly
            260                 265                 270
Gly Phe Thr Ala Asn Thr Ser Leu Ala His Tyr Cys Arg Asp Asn Gly
        275                 280                 285
Leu Leu Leu His Ile His Arg Ala Met His Ala Val Ile Asp Arg Gln
    290                 295                 300
Lys Asn His Gly Met His Phe Arg Val Leu Ala Lys Ala Leu Arg Met
305                 310                 315                 320
Ser Gly Gly Asp His Ile His Ala Gly Thr Val Val Gly Lys Leu Glu
                325                 330                 335
Gly Glu Arg Glu Met Thr Leu Gly Phe Val Asp Leu Leu Arg Asp Asp
            340                 345                 350
Phe Ile Glu Lys Asp Arg Ala Arg Gly Ile Phe Phe Thr Gln Asp Trp
        355                 360                 365
Val Ser Met Pro Gly Val Ile Pro Val Ala Ser Gly Gly Ile His Val
    370                 375                 380
Trp His Met Pro Ala Leu Thr Glu Ile Phe Gly Asp Asp Ser Val Leu
385                 390                 395                 400
Gln Phe Gly Gly Gly Thr Leu Gly His Pro Trp Gly Asn Ala Pro Gly
                405                 410                 415
Ala Ala Ala Asn Arg Val Ala Leu Glu Ala Cys Val Gln Ala Arg Asn
            420                 425                 430
Glu Gly Arg Asp Leu Ala Arg Glu Gly Asn Glu Ile Ile Arg Ser Ala
        435                 440                 445
Cys Lys Trp Ser Pro Glu Leu Ala Ala Ala Cys Glu Ile Trp Lys Ala
    450                 455                 460
Ile Lys Phe Glu Phe Glu Pro Val Asp Lys Leu Asp Ser
465                 470                 475

<210>  20
<211>  475
<212>  PRT
<213>  Glycine max

<400>  20

Met Ser Pro Gln Thr Glu Thr Lys Ala Ser Val Gly Phe Lys Ala Gly
1               5                   10                  15
Val Lys Asp Tyr Lys Leu Thr Tyr Tyr Thr Pro Asp Tyr Glu Thr Lys
            20                  25                  30
Asp Thr Asp Ile Leu Ala Ala Phe Arg Val Thr Pro Gln Pro Gly Val
        35                  40                  45
Pro Pro Glu Glu Ala Gly Ala Ala Val Ala Ala Glu Ser Ser Thr Gly
    50                  55                  60
Thr Trp Thr Thr Val Trp Thr Asp Gly Leu Thr Ser Leu Asp Arg Tyr
65                  70                  75                  80
```

```
Lys Gly Arg Cys Tyr Gly Leu Glu Pro Val Ala Gly Glu Glu Asn Gln
                85                  90                  95
Tyr Ile Ala Tyr Val Ala Tyr Pro Leu Asp Leu Phe Glu Glu Gly Ser
            100                 105                 110
Val Thr Asn Met Phe Thr Ser Ile Val Gly Asn Val Phe Gly Phe Lys
        115                 120                 125
Ala Leu Arg Ala Leu Arg Leu Glu Asp Leu Arg Ile Pro Thr Ala Tyr
    130                 135                 140
Ile Lys Thr Phe Gln Gly Pro Pro His Gly Ile Gln Val Glu Arg Asp
145                 150                 155                 160
Lys Leu Asn Lys Tyr Gly Arg Pro Leu Leu Gly Cys Thr Ile Lys Pro
                165                 170                 175
Lys Leu Gly Leu Ser Ala Lys Asn Tyr Gly Arg Ala Val Tyr Glu Cys
            180                 185                 190
Leu Arg Gly Gly Leu Asp Phe Thr Lys Asp Asp Glu Asn Val Asn Ser
        195                 200                 205
Gln Pro Phe Met Arg Trp Arg Asp Arg Phe Leu Phe Cys Ala Glu Ala
    210                 215                 220
Ile Phe Lys Ser Gln Ala Glu Thr Gly Glu Ile Lys Gly His Tyr Leu
225                 230                 235                 240
Asn Ala Thr Ala Gly Thr Cys Glu Glu Met Met Lys Arg Ala Val Phe
                245                 250                 255
Ala Arg Glu Leu Gly Val Pro Ile Val Met His Asp Tyr Leu Thr Gly
            260                 265                 270
Gly Phe Thr Ala Asn Thr Ser Leu Ala His Tyr Cys Arg Asp Asn Gly
        275                 280                 285
Leu Leu Leu His Ile His Arg Ala Met His Ala Val Ile Asp Arg Gln
    290                 295                 300
Lys Asn His Gly Met His Phe Arg Val Leu Ala Lys Ala Leu Arg Leu
305                 310                 315                 320
Ser Gly Gly Asp His Val His Ala Gly Thr Val Val Gly Lys Leu Glu
                325                 330                 335
Gly Glu Arg Glu Ile Thr Leu Gly Phe Val Asp Leu Leu Arg Asp Asp
            340                 345                 350
Phe Val Glu Lys Asp Arg Ser Arg Gly Ile Tyr Phe Thr Gln Asp Trp
        355                 360                 365
Val Ser Leu Pro Gly Val Leu Pro Val Ala Ser Gly Gly Ile His Val
    370                 375                 380
Trp His Met Pro Ala Leu Thr Glu Ile Phe Gly Asp Asp Ser Val Leu
385                 390                 395                 400
Gln Phe Gly Gly Gly Thr Leu Gly His Pro Trp Gly Asn Ala Pro Gly
                405                 410                 415
Ala Val Ala Asn Arg Val Ala Leu Glu Ala Cys Val Gln Ala Arg Asn
            420                 425                 430
Glu Gly Arg Asp Leu Ala Arg Glu Gly Asn Glu Ile Ile Arg Glu Ala
        435                 440                 445
Ser Lys Trp Ser Pro Glu Leu Ala Ala Ala Cys Glu Val Trp Lys Glu
    450                 455                 460
Ile Lys Phe Glu Phe Glu Ala Met Asp Thr Leu
465                 470                 475

<210>  21
<211>  476
<212>  PRT
<213>  Saccharum officinarum

<400>  21

Met Ser Pro Gln Thr Glu Thr Lys Ala Ser Val Gly Phe Lys Ala Gly
1               5                  10                  15
Val Lys Asp Tyr Lys Leu Thr Tyr Tyr Thr Pro Glu Tyr Glu Thr Lys
```

```
            20                  25                  30
Asp Thr Asp Ile Leu Ala Ala Phe Arg Val Thr Pro Gln Leu Gly Val
        35                  40                  45
Pro Pro Glu Glu Ala Gly Ala Ala Val Ala Ala Glu Ser Ser Thr Gly
    50                  55                  60
Thr Trp Thr Thr Val Trp Thr Asp Gly Leu Thr Ser Leu Asp Arg Tyr
65                  70                  75                  80
Lys Gly Arg Cys Tyr His Ile Glu Pro Val Pro Gly Asp Pro Asp Gln
                85                  90                  95
Tyr Ile Cys Tyr Val Ala Tyr Pro Leu Asp Leu Phe Glu Glu Gly Ser
            100                 105                 110
Val Thr Asn Met Phe Thr Ser Ile Val Gly Asn Val Phe Gly Phe Lys
        115                 120                 125
Ala Leu Arg Ala Leu Arg Leu Glu Asp Leu Arg Ile Pro Pro Ala Tyr
    130                 135                 140
Val Lys Thr Phe Gln Gly Pro Pro His Gly Ile Gln Val Glu Arg Asp
145                 150                 155                 160
Lys Leu Asn Lys Tyr Gly Arg Pro Leu Leu Gly Cys Thr Ile Lys Pro
                165                 170                 175
Lys Leu Gly Leu Ser Ala Lys Asn Tyr Gly Arg Ala Cys Tyr Glu Cys
            180                 185                 190
Leu Arg Gly Gly Leu Asp Phe Thr Lys Asp Asp Glu Asn Val Asn Ser
        195                 200                 205
Gln Pro Phe Met Arg Trp Arg Asp Arg Phe Val Phe Cys Ala Glu Ala
    210                 215                 220
Ile Tyr Lys Ala Gln Ala Glu Thr Gly Glu Ile Lys Gly His Tyr Leu
225                 230                 235                 240
Asn Ala Thr Ala Gly Thr Cys Glu Glu Met Ile Lys Arg Ala Val Phe
                245                 250                 255
Ala Lys Glu Leu Gly Val Pro Ile Val Met His Asp Tyr Leu Thr Gly
            260                 265                 270
Gly Phe Thr Ala Asn Thr Thr Leu Ser His Tyr Cys Arg Asp Asn Gly
        275                 280                 285
Leu Leu Leu His Ile His Arg Ala Met His Ala Val Ile Asp Arg Gln
    290                 295                 300
Lys Asn His Gly Met His Phe Arg Val Leu Ala Lys Ala Leu Arg Met
305                 310                 315                 320
Ser Gly Gly Asp His Ile His Ser Gly Thr Val Val Gly Lys Leu Glu
                325                 330                 335
Gly Glu Arg Glu Ile Thr Leu Gly Phe Val Asp Leu Leu Arg Asp Asp
            340                 345                 350
Phe Ile Glu Lys Asp Arg Ser Arg Gly Ile Phe Phe Thr Gln Asp Trp
        355                 360                 365
Val Ser Met Pro Gly Val Ile Pro Val Ala Ser Gly Gly Ile His Val
    370                 375                 380
Trp His Met Pro Ala Leu Thr Glu Ile Phe Gly Asp Asp Ser Val Leu
385                 390                 395                 400
Gln Phe Gly Gly Gly Thr Leu Gly His Pro Trp Gly Asn Ala Pro Gly
                405                 410                 415
Ala Ala Ala Asn Arg Val Ala Leu Glu Ala Cys Val Gln Ala Arg Asn
            420                 425                 430
Glu Gly Arg Asp Leu Ala Arg Glu Gly Asn Glu Ile Ile Lys Ala Ala
        435                 440                 445
Cys Lys Trp Ser Ala Glu Leu Ala Ala Ala Cys Glu Ile Trp Lys Glu
    450                 455                 460
Ile Lys Phe Asp Thr Phe Lys Ala Met Asp Thr Leu
465                 470                 475
```

<210> 22
<211> 1845
<212> DNA

<213> Synechococcus PCC6301

<400> 22

```
atggccaaga cgcaatctgc cgcaggctat aaggccgggg tgaaggacta caaactcacc     60
tattacaccc ccgattacac ccccaaagac actgacctgc tggcggcttt ccgcttcagc    120
cctcagccgg gtgtccctgc tgacgaagct ggtgcggcga tcgcggctga atcttcgacc    180
ggtacctgga ccaccgtgtg gaccgacttg ctgaccgaca tggatcggta caaaggcaag    240
tgctaccaca tcgagccggt gcaaggcgaa gagaactcct actttgcgtt catcgcttac    300
ccgctcgacc tgtttgaaga agggtcggtc accaacatcc tgacctcgat cgtcggtaac    360
gtgtttggct tcaaagctat ccgttcgctg cgtctggaag acatccgctt ccccgtcgcc    420
ttggtcaaaa ccttccaagg tcctccccac ggtatccaag tcgagcgcga cctgctgaac    480
aagtacggcc gtccgatgct gggttgcacg atcaaaccaa aactcggtct gtcggcgaaa    540
aactacggtc gtgccgtcta cgaatgtctg cgcggcggtc tggacttcac caaagacgac    600
gaaaacatca actcgcagcc gttccaacgc tggcgcgatc gcttcctgtt tgtggctgat    660
gcaatccaca aatcgcaagc agaaaccggt gaaatcaaag gtcactacct gaacgtgacc    720
gcgccgacct gcgaagaaat gatgaaacgg gctgagttcg ctaaagaact cggcatgccg    780
atcatcatgc atgacttctt gacggctggt ttcaccgcca acaccacctt ggcaaaatgg    840
tgccgcgaca acggcgtcct gctgcacatc caccgtgcaa tgcacgcggt gatcgaccgt    900
cagcgtaacc acgggattca cttccgtgtc ttggccaagt gtttgcgtct gtccggtggt    960
gaccacctcc actccggcac cgtcgtcggc aaactggaag gcgacaaagc ttcgaccttg   1020
ggctttgttg acttgatgcg cgaagaccac atcgaagctg accgcagccg tggggtcttc   1080
ttcacccaag attgggcgtc gatgccgggc gtgctgccgg ttgcttccgg tggtatccac   1140
gtgtggcaca tgcccgcact ggtggaaatc ttcggtgatg actccgttct ccagttcggt   1200
ggcggcacct tgggtcaccc ctggggtaat gctcctggtg caaccgcgaa ccgtgttgcc   1260
ttggaagctt gcgtccaagc tcggaacgaa ggtcgcgacc tctaccgtga aggcggcgac   1320
atccttcgtg aagctggcaa gtggtcgcct gaactggctg ctgccctcga cctctggaaa   1380
gagatcaagt tcgaattcga aacgatggac aagctctaag gagcctctga ctatcgctgg   1440
gggagtgagc gttgctgcgt aaagctttct ccccagcctt tcgacttaac ctttcaggat   1500
ttctgaatca tgagcatgaa aactctgccc aaagagcgtc gtttcgagac tttctcgtac   1560
ctgcctcccc tcagcgatcg ccaaatcgct gcacaaatcg agtacatgat cgagcaaggc   1620
ttccaccct tgatcgagtt caacgagcac tcgaatccgg aagagttcta ctggacgatg   1680
tggaagctcc ccctgtttga ctgcaagagc cctcagcaag tcctcgatga agtgcgtgag   1740
tgccgcagcg aatacggtga ttgctacatc cgtgtcgctg gcttcgacaa catcaagcag   1800
tgccaaaccg tgagcttcat cgttcatcgt cccggccgat actaa                   1845
```

<210> 23
<211> 111
<212> PRT
<213> Synechococcus PCC6301

<400> 23

```
Met Ser Met Lys Thr Leu Pro Lys Glu Arg Arg Phe Glu Thr Phe Ser
1               5                  10                  15
Tyr Leu Pro Pro Leu Ser Asp Arg Gln Ile Ala Ala Gln Ile Glu Tyr
            20                  25                  30
Met Ile Glu Gln Gly Phe His Pro Leu Ile Glu Phe Asn Glu His Ser
        35                  40                  45
Asn Pro Glu Glu Phe Tyr Trp Thr Met Trp Lys Leu Pro Leu Phe Asp
    50                  55                  60
Cys Lys Ser Pro Gln Gln Val Leu Asp Glu Val Arg Glu Cys Arg Ser
65                  70                  75                  80
Glu Tyr Gly Asp Cys Tyr Ile Arg Val Ala Gly Phe Asp Asn Ile Lys
                85                  90                  95
Gln Cys Gln Thr Val Ser Phe Ile Val His Arg Pro Gly Arg Tyr
            100                 105                 110
```

<210> 24
<211> 472
<212> PRT
<213> Synechococcus PCC6301

<400> 24

```
Met Ala Lys Thr Gln Ser Ala Ala Gly Tyr Lys Ala Gly Val Lys Asp
1               5                  10                  15
Tyr Lys Leu Thr Tyr Tyr Thr Pro Asp Tyr Thr Pro Lys Asp Thr Asp
            20                  25                  30
Leu Leu Ala Ala Phe Arg Phe Ser Pro Gln Pro Gly Val Pro Ala Asp
        35                  40                  45
Glu Ala Gly Ala Ala Ile Ala Ala Glu Ser Ser Thr Gly Thr Trp Thr
    50                  55                  60
Thr Val Trp Thr Asp Leu Leu Thr Asp Met Asp Arg Tyr Lys Gly Lys
65                  70                  75                  80
Cys Tyr His Ile Glu Pro Val Gln Gly Glu Glu Asn Ser Tyr Phe Ala
                85                  90                  95
Phe Ile Ala Tyr Pro Leu Asp Leu Phe Glu Glu Gly Ser Val Thr Asn
            100                 105                 110
Ile Leu Thr Ser Ile Val Gly Asn Val Phe Gly Phe Lys Ala Ile Arg
        115                 120                 125
Ser Leu Arg Leu Glu Asp Ile Arg Phe Pro Val Ala Leu Val Lys Thr
    130                 135                 140
Phe Gln Gly Pro Pro His Gly Ile Gln Val Glu Arg Asp Leu Leu Asn
145                 150                 155                 160
Lys Tyr Gly Arg Pro Met Leu Gly Cys Thr Ile Lys Pro Lys Leu Gly
                165                 170                 175
Leu Ser Ala Lys Asn Tyr Gly Arg Ala Val Tyr Glu Cys Leu Arg Gly
            180                 185                 190
Gly Leu Asp Phe Thr Lys Asp Asp Glu Asn Ile Asn Ser Gln Pro Phe
        195                 200                 205
Gln Arg Trp Arg Asp Arg Phe Leu Phe Val Ala Asp Ala Ile His Lys
    210                 215                 220
Ser Gln Ala Glu Thr Gly Glu Ile Lys Gly His Tyr Leu Asn Val Thr
225                 230                 235                 240
Ala Pro Thr Cys Glu Glu Met Met Lys Arg Ala Glu Phe Ala Lys Glu
                245                 250                 255
Leu Gly Met Pro Ile Ile Met His Asp Phe Leu Thr Ala Gly Phe Thr
            260                 265                 270
Ala Asn Thr Thr Leu Ala Lys Trp Cys Arg Asp Asn Gly Val Leu Leu
        275                 280                 285
His Ile His Arg Ala Met His Ala Val Ile Asp Arg Gln Arg Asn His
    290                 295                 300
Gly Ile His Phe Arg Val Leu Ala Lys Cys Leu Arg Leu Ser Gly Gly
305                 310                 315                 320
Asp His Leu His Ser Gly Thr Val Val Gly Lys Leu Glu Gly Asp Lys
                325                 330                 335
Ala Ser Thr Leu Gly Phe Val Asp Leu Met Arg Glu Asp His Ile Glu
            340                 345                 350
Ala Asp Arg Ser Arg Gly Val Phe Phe Thr Gln Asp Trp Ala Ser Met
        355                 360                 365
Pro Gly Val Leu Pro Val Ala Ser Gly Gly Ile His Val Trp His Met
    370                 375                 380
Pro Ala Leu Val Glu Ile Phe Gly Asp Asp Ser Val Leu Gln Phe Gly
385                 390                 395                 400
Gly Gly Thr Leu Gly His Pro Trp Gly Asn Ala Pro Gly Ala Thr Ala
                405                 410                 415
Asn Arg Val Ala Leu Glu Ala Cys Val Gln Ala Arg Asn Glu Gly Arg
            420                 425                 430
Asp Leu Tyr Arg Glu Gly Gly Asp Ile Leu Arg Glu Ala Gly Lys Trp
        435                 440                 445
Ser Pro Glu Leu Ala Ala Ala Leu Asp Leu Trp Lys Glu Ile Lys Phe
    450                 455                 460
```

```
Glu Phe Glu Thr Met Asp Lys Leu
465                 470
```

<210> 25
<211> 1845
<212> DNA
<213> Artificial

<220>
<223> Com base no nucleotídeo PCC6301 rbcLS de Synechococcus Sp. do tipo selvagem

<400> 25

```
atggccaaga cgcaatctgc cgcaggctat aaggccgggg tgaaggacta caaactcacc     60
tattggaccc ccgattacac ccccaaagac actgacctgc tggcggcttt ccgcttcagc    120
cctcagccgg gtgtccctgc tatcgaagct ggtgcggcga tcgcggctga atcttcgacc    180
ggtacctgga ccaccgtgtg gaccgacttg ctgaccgaca tggatcggta caaaggcaag    240
tgctaccaca tcgagccggt gcaaggcgaa gagaactcct actttgcgtt catcgcttac    300
ccgctcgacc tgtttgaaga agggtcggtc accaacatcc tgacctcgat cgtcggtaac    360
gtgtttggct tcaaagctat ccgttcgctg cgtctggaag acatccgctt ccccgtcgcc    420
ttggtcaaaa ccttccaagg tcctccccac ggtatccaag tcgagcgcga cctgctgaac    480
aagtacggcc gtccgatgct gggttgcacg atcaaaccaa aactcggtct gtcggcgaaa    540
aactacggtc gtgccgtcta cgaatgtctg cgcggcggtc tggacttcac caaagacgac    600
gaaaacatca actcgcagcc gttccaacgc tggcgcgatc gcttcctgtt tgtggctgat    660
gcaatccaca aatcgcaagc agaaaccggt gaaatcaaag gtcactacct gaacgtgacc    720
gcgccgacct gcgaagaaat gatgaaacgg gctgagttcg ctaaagaact cggcatgccg    780
atcatcatgc atgacttctt gacggctggt ttcaccgcca acaccacctt ggcaaaatgg    840
tgccgcgaca acggcgtcct gctgcacatc caccgtgcaa tgcacgcggt gatcgaccgt    900
cagcgtaacc acgggattca cttccgtgtc ttggccaagt gtttgcgtct gtccggtggt    960
gaccacctcc actccggcac cgtcgtcggc aaactggaag gcgacaaagc ttcgaccttg   1020
ggctttgttg acttgatgcg cgaagaccac atcgaagctg accgcagccg tggggtcttc   1080
ttcacccaag attgggcgtc gatgccgggc gtgctgccgg ttgcttccgg tggtatccac   1140
gtgtggcaca tgcccgcact ggtggaaatc ttcggtgatg actccgttct ccagttcggt   1200
ggcggcacct tgggtcaccc ctggggtaat gctcctggtg caaccgcgaa ccgtgttgcc   1260
ttggaagctt gcgtccaagc tcggaacgaa ggtcgcgacc tctaccgtga aggcggcgac   1320
atccttcgtg aagctggcaa gtggtcgcct gaactggctg ctgccctcga cctctggaaa   1380
gagatcaagt tcgaattcga aacgatggac aagctctaag gagcctctga ctatcgctgg   1440
gggagtgagc gttgctgcgt aaagctttct ccccagcctt tcgacttaac ctttcaggat   1500
ttctgaatca tgagcatgaa aactctgccc aaagagcgtc gtttcgagac tttctcgtac   1560
ctgcctcccc tcagcgatcg ccaaatcgct gcacaaatcg agtacatgat cgagcaaggc   1620
ttccaccccc tgatcgagtt caacgagcac tcgaatccgg aagagttcta ctggacgatg   1680
tggaagctcc ccctgtttga ctgcaagagc cctcagcaag tcctcgatga agtgcgtgag   1740
tgccgcagcg aatacggtga ttgctacatc cgtgtcgctg gcttcgacaa catcaagcag   1800
tgccaaaccg tgagcttcat cgttcatcgt cccggccgat actaa                   1845
```

<210> 26
<211> 472
<212> PRT
<213> Artificial

<220>
<223> Com base na sequência traduzida de PCC6301 rbcLS de Synechococcus Sp. do tipo selvagem

<400> 26

```
Met Ala Lys Thr Gln Ser Ala Ala Gly Tyr Lys Ala Gly Val Lys Asp
1               5                  10                  15
Tyr Lys Leu Thr Tyr Trp Thr Pro Asp Tyr Thr Pro Lys Asp Thr Asp
            20                  25                  30
Leu Leu Ala Ala Phe Arg Phe Ser Pro Gln Pro Gly Val Pro Ala Ile
```

```
        35                  40                  45
Glu Ala Gly Ala Ala Ile Ala Ala Glu Ser Ser Thr Gly Thr Trp Thr
    50                  55                  60
Thr Val Trp Thr Asp Leu Leu Thr Asp Met Asp Arg Tyr Lys Gly Lys
65                  70                  75                  80
Cys Tyr His Ile Glu Pro Val Gln Gly Glu Glu Asn Ser Tyr Phe Ala
                85                  90                  95
Phe Ile Ala Tyr Pro Leu Asp Leu Phe Glu Glu Gly Ser Val Thr Asn
            100                 105                 110
Ile Leu Thr Ser Ile Val Gly Asn Val Phe Gly Phe Lys Ala Ile Arg
        115                 120                 125
Ser Leu Arg Leu Glu Asp Ile Arg Phe Pro Val Ala Leu Val Lys Thr
    130                 135                 140
Phe Gln Gly Pro Pro His Gly Ile Gln Val Glu Arg Asp Leu Leu Asn
145                 150                 155                 160
Lys Tyr Gly Arg Pro Met Leu Gly Cys Thr Ile Lys Pro Lys Leu Gly
                165                 170                 175
Leu Ser Ala Lys Asn Tyr Gly Arg Ala Val Tyr Glu Cys Leu Arg Gly
            180                 185                 190
Gly Leu Asp Phe Thr Lys Asp Asp Glu Asn Ile Asn Ser Gln Pro Phe
        195                 200                 205
Gln Arg Trp Arg Asp Arg Phe Leu Phe Val Ala Asp Ala Ile His Lys
    210                 215                 220
Ser Gln Ala Glu Thr Gly Glu Ile Lys Gly His Tyr Leu Asn Val Thr
225                 230                 235                 240
Ala Pro Thr Cys Glu Glu Met Met Lys Arg Ala Glu Phe Ala Lys Glu
                245                 250                 255
Leu Gly Met Pro Ile Ile Met His Asp Phe Leu Thr Ala Gly Phe Thr
            260                 265                 270
Ala Asn Thr Thr Leu Ala Lys Trp Cys Arg Asp Asn Gly Val Leu Leu
        275                 280                 285
His Ile His Arg Ala Met His Ala Val Ile Asp Arg Gln Arg Asn His
    290                 295                 300
Gly Ile His Phe Arg Val Leu Ala Lys Cys Leu Arg Leu Ser Gly Gly
305                 310                 315                 320
Asp His Leu His Ser Gly Thr Val Val Gly Lys Leu Glu Gly Asp Lys
                325                 330                 335
Ala Ser Thr Leu Gly Phe Val Asp Leu Met Arg Glu Asp His Ile Glu
            340                 345                 350
Ala Asp Arg Ser Arg Gly Val Phe Phe Thr Gln Asp Trp Ala Ser Met
        355                 360                 365
Pro Gly Val Leu Pro Val Ala Ser Gly Gly Ile His Val Trp His Met
    370                 375                 380
Pro Ala Leu Val Glu Ile Phe Gly Asp Asp Ser Val Leu Gln Phe Gly
385                 390                 395                 400
Gly Gly Thr Leu Gly His Pro Trp Gly Asn Ala Pro Gly Ala Thr Ala
                405                 410                 415
Asn Arg Val Ala Leu Glu Ala Cys Val Gln Ala Arg Asn Glu Gly Arg
            420                 425                 430
Asp Leu Tyr Arg Glu Gly Gly Asp Ile Leu Arg Glu Ala Gly Lys Trp
        435                 440                 445
Ser Pro Glu Leu Ala Ala Ala Leu Asp Leu Trp Lys Glu Ile Lys Phe
    450                 455                 460
Glu Phe Glu Thr Met Asp Lys Leu
465                 470
```

<210> 27
<211> 1845
<212> DNA
<213> Artificial

<220>
<223> Com base no nucleotídeo PCC6301 rbcLS de Synechococcus Sp. do tipo selvagem

<400> 27
atggccaaga cgcaatctgc cgcaggctat aaggccgggg tgaaggacta caaactcacc 60
tattggaccc ccgattacac ccccaaagac actgacctgc tggcggcttt ccgcttcagc 120
cctcagccgg gtgtccctgc tgtcgaagct ggtgcggcga tcgcggctga atcttcgacc 180
ggtacctgga ccaccgtgtg gaccgacttg ctgaccgaca tggatcggta caaaggcaag 240
tgctaccaca tcgagccggt gcaaggcgaa gagaactcct actttgcgtt catcgcttac 300
ccgctcgacc tgtttgaaga agggtcggtc accaacatcc tgacctcgat cgtcggtaac 360
gtgtttggct tcaaagctat ccgttcgctg cgtctggaag acatccgctt ccccgtcgcc 420
ttggtcaaaa ccttccaagg tcctccccac ggtatccaag tcgagcgcga cctgctgaac 480
aagtacggcc gtccgatgct gggttgcacg atcaaaccaa aactcggtct gtcggcgaaa 540
aactacggtc gtgccgtcta cgaatgtctg cgcggcggtc tggacttcac caaagacgac 600
gaaaacatca actcgcagcc gttccaacgc tggcgcgatc gcttcctgtt tgtggctgat 660
gcaatccaca aatcgcaagc agaaaccggt gaaatcaaag gtcactacct gaacgtgacc 720
gcgccgacct gcgaagaaat gatgaaacgg gctgagttcg ctaaagaact cggcatgccg 780
atcatcatgc atgacttctt gacggctggt ttcaccgcca acaccacctt ggcaaaatgg 840
tgccgcgaca acggcgtcct gctgcacatc caccgtgcaa tgcacgcggt gatcgaccgt 900
cagcgtaacc acgggattca cttccgtgtc ttggccaagt gtttgcgtct gtccggtggt 960
gaccacctcc actccggcac cgtcgtcggc aaactggaag gcgacaaagc ttcgaccttg 1020
ggctttgttg acttgatgcg cgaagaccac atcgaagctg accgcagccg tggggtcttc 1080
ttcacccaag attgggcgtc gatgccgggc gtgctgccgg ttgcttccgg tggtatccac 1140
gtgtggcaca tgcccgcact ggtggaaatc ttcggtgatg actccgttct ccagttcggt 1200
ggcggcacct tgggtcaccc ctggggtaat gctcctggtg caaccgcgaa ccgtgttgcc 1260
ttggaagctt gcgtccaagc tcggaacgaa ggtcgcgacc tctaccgtga aggcggcgac 1320
atccttcgtg aagctggcaa gtggtcgcct gaactggctg ctgccctcga cctctggaaa 1380
gagatcaagt tcgaattcga aacgatggac aagctctaag gagcctctga ctatcgctgg 1440
gggagtgagc gttgctgcgt aaagctttct ccccagcctt tcgacttaac ctttcaggat 1500
ttctgaatca tgagcatgaa aactctgccc aaagagcgtc gtttcgagac tttctcgtac 1560
ctgcctcccc tcagcgatcg ccaaatcgct gcacaaatcg agtacatgat cgagcaaggc 1620
ttccaccccc tgatcgagtt caacgagcac tcgaatccgg aagagttcta ctggacgatg 1680
tggaagctcc ccctgtttga ctgcaagagc cctcagcaag tcctcgatga agtgcgtgag 1740
tgccgcagcg aatacggtga ttgctacatc cgtgtcgctg gcttcgacaa catcaagcag 1800
tgccaaaccg tgagcttcat cgttcatcgt cccggccgat actaa 1845

<210> 28
<211> 472
<212> PRT
<213> Artificial

<220>
<223> Com base na sequência traduzida de PCC6301 rbcLS de Synechococcus Sp. do tipo selvagem

<400> 28

Met Ala Lys Thr Gln Ser Ala Ala Gly Tyr Lys Ala Gly Val Lys Asp
1 5 10 15
Tyr Lys Leu Thr Tyr Trp Thr Pro Asp Tyr Thr Pro Lys Asp Thr Asp
20 25 30
Leu Leu Ala Ala Phe Arg Phe Ser Pro Gln Pro Gly Val Pro Ala Val
35 40 45
Glu Ala Gly Ala Ala Ile Ala Ala Glu Ser Ser Thr Gly Thr Trp Thr
50 55 60
Thr Val Trp Thr Asp Leu Leu Thr Asp Met Asp Arg Tyr Lys Gly Lys
65 70 75 80
Cys Tyr His Ile Glu Pro Val Gln Gly Glu Glu Asn Ser Tyr Phe Ala
85 90 95
Phe Ile Ala Tyr Pro Leu Asp Leu Phe Glu Glu Gly Ser Val Thr Asn

```
            100                 105                 110
Ile Leu Thr Ser Ile Val Gly Asn Val Phe Gly Phe Lys Ala Ile Arg
        115                 120                 125
Ser Leu Arg Leu Glu Asp Ile Arg Phe Pro Val Ala Leu Val Lys Thr
    130                 135                 140
Phe Gln Gly Pro Pro His Gly Ile Gln Val Glu Arg Asp Leu Leu Asn
145                 150                 155                 160
Lys Tyr Gly Arg Pro Met Leu Gly Cys Thr Ile Lys Pro Lys Leu Gly
                165                 170                 175
Leu Ser Ala Lys Asn Tyr Gly Arg Ala Val Tyr Glu Cys Leu Arg Gly
            180                 185                 190
Gly Leu Asp Phe Thr Lys Asp Asp Glu Asn Ile Asn Ser Gln Pro Phe
        195                 200                 205
Gln Arg Trp Arg Asp Arg Phe Leu Phe Val Ala Asp Ala Ile His Lys
    210                 215                 220
Ser Gln Ala Glu Thr Gly Glu Ile Lys Gly His Tyr Leu Asn Val Thr
225                 230                 235                 240
Ala Pro Thr Cys Glu Glu Met Met Lys Arg Ala Glu Phe Ala Lys Glu
                245                 250                 255
Leu Gly Met Pro Ile Ile Met His Asp Phe Leu Thr Ala Gly Phe Thr
            260                 265                 270
Ala Asn Thr Thr Leu Ala Lys Trp Cys Arg Asp Asn Gly Val Leu Leu
        275                 280                 285
His Ile His Arg Ala Met His Ala Val Ile Asp Arg Gln Arg Asn His
    290                 295                 300
Gly Ile His Phe Arg Val Leu Ala Lys Cys Leu Arg Leu Ser Gly Gly
305                 310                 315                 320
Asp His Leu His Ser Gly Thr Val Val Gly Lys Leu Glu Gly Asp Lys
                325                 330                 335
Ala Ser Thr Leu Gly Phe Val Asp Leu Met Arg Glu Asp His Ile Glu
            340                 345                 350
Ala Asp Arg Ser Arg Gly Val Phe Phe Thr Gln Asp Trp Ala Ser Met
        355                 360                 365
Pro Gly Val Leu Pro Val Ala Ser Gly Gly Ile His Val Trp His Met
    370                 375                 380
Pro Ala Leu Val Glu Ile Phe Gly Asp Asp Ser Val Leu Gln Phe Gly
385                 390                 395                 400
Gly Gly Thr Leu Gly His Pro Trp Gly Asn Ala Pro Gly Ala Thr Ala
                405                 410                 415
Asn Arg Val Ala Leu Glu Ala Cys Val Gln Ala Arg Asn Glu Gly Arg
            420                 425                 430
Asp Leu Tyr Arg Glu Gly Gly Asp Ile Leu Arg Glu Ala Gly Lys Trp
        435                 440                 445
Ser Pro Glu Leu Ala Ala Ala Leu Asp Leu Trp Lys Glu Ile Lys Phe
    450                 455                 460
Glu Phe Glu Thr Met Asp Lys Leu
465                 470
```

<210> 29
<211> 1845
<212> DNA
<213> Artificial

<220>
<223> Com base no nucleotídeo PCC6301 rbcLS de Synechococcus Sp. do tipo selvagem

<400> 29

```
atggccaaga cgcaatctgc cgcaggctat aaggccgggg tgaaggacta caaactcggc     60
tattacaccc ccgattacac ccccaaagac actgacctgc tggcggcttt ccgcttcagc    120
cctcagccgg gtgtccctgc tgacgaagct ggtgcggcga tcgcggctga atcttcgacc    180
```

```
ggtacctgga ccaccgtgtg gaccgacttg ctgaccgaca tggatcggta ccgcggcaag    240
tgctaccaca tcgagccggt gcaaggcgaa gagaactcct actttgcgtt catcgcttac    300
ccgctcgacc tgtttgaaga agggtcggtc accaacatcc tgacctcgat cgtcggtaac    360
gtgtttggct tcaaagctat ccgttcgctg cgtctggaag acatccgctt ccccgtcgcc    420
ttggtcaaaa ccttccaagg tcctccccac ggtatccaag tcgagcgcga cctgctgaac    480
aagtacggcc gtccgatgct gggttgcacg atcaaaccaa aactcggtct gtcggcgaaa    540
aactacggtc gtgccgtcta cgaatgtctg cgcggcggtc tggacttcac caaagacgac    600
gaaaacatca actcgcagcc gttccaacgc tggcgcgatc gcttcctgtt tgtggctgat    660
gcaatccaca aatcgcaagc agaaaccggt gaaatcaaag gtcactacct gaacgtgacc    720
gcgccgacct gcgaagaaat gatgaaacgg gctgagttcg ctaaagaact cggcatgccg    780
atcatcatgc atgacttctt gacggctggt ttcaccgcca acaccacctt ggcaaaatgg    840
tgccgcgaca acggcgtcct gctgcacatc caccgtgcaa tgcacgcggt gatcgaccgt    900
cagcgtaacc acgggattca cttccgtgtc ttggccaagt gtttgcgtct gtccggtggt    960
gaccacctcc actccggcac cgtcgtcggc aaactggaag gcgacaaagc ttcgaccttg   1020
ggctttgttg acttgatgcg cgaagaccac atcgaagctg accgcagccg tggggtcttc   1080
ttcacccaag attgggcgtc gatgccgggc gtgctgccgg ttgcttccgg tggtatccac   1140
gtgtggcaca tgcccgcact ggtggaaatc ttcggtgatg actccgttct ccagttcggt   1200
ggcggcacct tgggtcaccc ctggggtaat gctcctggtg caaccgcgaa ccgtgttgcc   1260
ttggaagctt gcgtccaagc tcggaacgaa ggtcgcgacc tctaccgtga aggcggcgac   1320
atccttcgtg aagctggcaa gtggtcgcct gaactggctg ctgccctcga cctctggaaa   1380
gagatcaagt tcgaattcga aacgatggac aagctctaag gagcctctga ctatcgctgg   1440
gggagtgagc gttgctgcgt aaagctttct ccccagcctt tcgacttaac ctttcaggat   1500
ttctgaatca tgagcatgaa aactctgccc aaagagcgtc gtttcgagac tttctcgtac   1560
ctgcctcccc tcagcgatcg ccaaatcgct gcacaaatcg agtacatgat cgagcaaggc   1620
ttccaccccc tgatcgagtt caacgagcac tcgaatccgg aagagttcta ctggacgatg   1680
tggaagctcc ccctgtttga ctgcaagagc cctcagcaag tcctcgatga agtgcgtgag   1740
tgccgcagcg aatacggtga ttgctacatc cgtgtcgctg gcttcgacaa catcaagcag   1800
tgccaaaccg tgagcttcat cgttcatcgt cccggccgat actaa                   1845
```

<210> 30
<211> 472
<212> PRT
<213> Artificial

<220>
<223> Com base na sequência traduzida de PCC6301 rbcLS de Synechococcus Sp. do tipo selvagem

<400> 30

```
Met Ala Lys Thr Gln Ser Ala Ala Gly Tyr Lys Ala Gly Val Lys Asp
1               5                  10                  15
Tyr Lys Leu Gly Tyr Tyr Thr Pro Asp Tyr Thr Pro Lys Asp Thr Asp
            20                  25                  30
Leu Leu Ala Ala Phe Arg Phe Ser Pro Gln Pro Gly Val Pro Ala Asp
        35                  40                  45
Glu Ala Gly Ala Ala Ile Ala Ala Glu Ser Ser Thr Gly Thr Trp Thr
    50                  55                  60
Thr Val Trp Thr Asp Leu Leu Thr Asp Met Asp Arg Tyr Arg Gly Lys
65                  70                  75                  80
Cys Tyr His Ile Glu Pro Val Gln Gly Glu Glu Asn Ser Tyr Phe Ala
                85                  90                  95
Phe Ile Ala Tyr Pro Leu Asp Leu Phe Glu Glu Gly Ser Val Thr Asn
            100                 105                 110
Ile Leu Thr Ser Ile Val Gly Asn Val Phe Gly Phe Lys Ala Ile Arg
        115                 120                 125
Ser Leu Arg Leu Glu Asp Ile Arg Phe Pro Val Ala Leu Val Lys Thr
    130                 135                 140
Phe Gln Gly Pro Pro His Gly Ile Gln Val Glu Arg Asp Leu Leu Asn
145                 150                 155                 160
Lys Tyr Gly Arg Pro Met Leu Gly Cys Thr Ile Lys Pro Lys Leu Gly
```

```
                165                 170                 175
Leu Ser Ala Lys Asn Tyr Gly Arg Ala Val Tyr Glu Cys Leu Arg Gly
            180                 185                 190
Gly Leu Asp Phe Thr Lys Asp Asp Glu Asn Ile Asn Ser Gln Pro Phe
        195                 200                 205
Gln Arg Trp Arg Asp Arg Phe Leu Phe Val Ala Asp Ala Ile His Lys
    210                 215                 220
Ser Gln Ala Glu Thr Gly Glu Ile Lys Gly His Tyr Leu Asn Val Thr
225                 230                 235                 240
Ala Pro Thr Cys Glu Glu Met Met Lys Arg Ala Glu Phe Ala Lys Glu
                245                 250                 255
Leu Gly Met Pro Ile Ile Met His Asp Phe Leu Thr Ala Gly Phe Thr
            260                 265                 270
Ala Asn Thr Thr Leu Ala Lys Trp Cys Arg Asp Asn Gly Val Leu Leu
        275                 280                 285
His Ile His Arg Ala Met His Ala Val Ile Asp Arg Gln Arg Asn His
    290                 295                 300
Gly Ile His Phe Arg Val Leu Ala Lys Cys Leu Arg Leu Ser Gly Gly
305                 310                 315                 320
Asp His Leu His Ser Gly Thr Val Val Gly Lys Leu Glu Gly Asp Lys
                325                 330                 335
Ala Ser Thr Leu Gly Phe Val Asp Leu Met Arg Glu Asp His Ile Glu
            340                 345                 350
Ala Asp Arg Ser Arg Gly Val Phe Phe Thr Gln Asp Trp Ala Ser Met
        355                 360                 365
Pro Gly Val Leu Pro Val Ala Ser Gly Gly Ile His Val Trp His Met
    370                 375                 380
Pro Ala Leu Val Glu Ile Phe Gly Asp Asp Ser Val Leu Gln Phe Gly
385                 390                 395                 400
Gly Gly Thr Leu Gly His Pro Trp Gly Asn Ala Pro Gly Ala Thr Ala
                405                 410                 415
Asn Arg Val Ala Leu Glu Ala Cys Val Gln Ala Arg Asn Glu Gly Arg
            420                 425                 430
Asp Leu Tyr Arg Glu Gly Gly Asp Ile Leu Arg Glu Ala Gly Lys Trp
        435                 440                 445
Ser Pro Glu Leu Ala Ala Ala Leu Asp Leu Trp Lys Glu Ile Lys Phe
    450                 455                 460
Glu Phe Glu Thr Met Asp Lys Leu
465                 470
```

<210> 31
<211> 1845
<212> DNA
<213> Artificial

<220>
<223> Com base no nucleotídeo PCC6301 rbcLS de Synechococcus Sp. do tipo selvagem

<400> 31

```
atggccaaga cgcaatctgc cgcaggctat aaggccgggg tgaaggacta caaactcacc     60
tattacaccc ccgattacac ccccaaagac actgacctgc tggcggcttt ccgcttcagc    120
cctcagccgg gtgtccctgc tgacgaagct gcggcggcga tcgcggctga atcttcgacc    180
ggtacctgga ccaccgtgtg gaccgacttg ctgaccgaca tggatcggta caaaggcaag    240
gcctaccacg tggagccggt gcaaggcgaa gagaactcct actttgcgtt catcgcttac    300
ccgctcgacc tgtttgaaga agggtcggtc accaacatcc tgacctcgat cgtcggtaac    360
gtgtttggct tcaaagctat ccgttcgctg cgtctggaag acatccgctt ccccgtcgcc    420
ttggtcaaaa ccttccaagg tcctccccac ggtatccaag tcgagcgcga cctgctgaac    480
aagtacggcc gtccgatgct gggttgcacg atcaaaccaa aactcggtct gtcggcgaaa    540
aactacggtc gtgccgtcta cgaatgtctg cgcggcggtc tggacttcac caaagacgac    600
gaaaacatca actcgcagcc gttccaacgc tggcgcgatc gcttcctgtt tgtggctgat    660
```

```
gcaatccaca aatcgcaagc agaaaccggt gaaatcaaag gtcactacct gaacgtgacc    720
gcgccgacct gcgaagaaat gatgaaacgg gctgagttcg ctaaagaact cggcatgccg    780
atcatcatgc atgacttctt gacggctggt ttcaccgcca acaccacctt ggcaaaatgg    840
tgccgcgaca acggcgtcct gctgcacatc caccgtgcaa tgcacgcggt gatcgaccgt    900
cagcgtaacc acgggattca cttccgtgtc ttggccaagt gtttgcgtct gtccggtggt    960
gaccacctcc actccggcac cgtcgtcggc aaactggaag gcgacaaagc ttcgaccttg   1020
ggctttgttg acttgatgcg cgaagaccac atcgaagctg accgcagccg tggggtcttc   1080
ttcacccaag attgggcgtc gatgccgggc gtgctgccgg ttgcttccgg tggtatccac   1140
gtgtggcaca tgcccgcact ggtggaaatc ttcggtgatg actccgttct ccagttcggt   1200
ggcggcacct tgggtcaccc ctggggtaat gctcctggtg caaccgcgaa ccgtgttgcc   1260
ttggaagctt gcgtccaagc tcggaacgaa ggtcgcgacc tctaccgtga aggcggcgac   1320
atccttcgtg aagctggcaa gtggtcgcct gaactggctg ctgccctcga cctctggaaa   1380
gagatcaagt tcgaattcga aacgatggac aagctctaag gagcctctga ctatcgctgg   1440
gggagtgagc gttgctgcgt aaagctttct ccccagcctt tcgacttaac ctttcaggat   1500
ttctgaatca tgagcatgaa aactctgccc aaagagcgtc gtttcgagac tttctcgtac   1560
ctgcctcccc tcagcgatcg ccaaatcgct gcacaaatcg agtacatgat cgagcaaggc   1620
ttccacccct tgatcgagtt caacgagcac tcgaatccgg aagagttcta ctggacgatg   1680
tggaagctcc ccctgtttga ctgcaagagc cctcagcaag tcctcgatga agtgcgtgag   1740
tgccgcagcg aatacggtga ttgctacatc cgtgtcgctg gcttcgacaa catcaagcag   1800
tgccaaaccg tgagcttcat cgttcatcgt cccggccgat actaa                   1845
```

<210> 32
<211> 472
<212> PRT
<213> Artificial

<220>
<223> Com base na sequência traduzida de PCC6301 rbcLS de Synechococcus Sp. do tipo selvagem

<400> 32

```
Met Ala Lys Thr Gln Ser Ala Ala Gly Tyr Lys Ala Gly Val Lys Asp
1               5                  10                  15
Tyr Lys Leu Thr Tyr Tyr Thr Pro Asp Tyr Thr Pro Lys Asp Thr Asp
            20                  25                  30
Leu Leu Ala Ala Phe Arg Phe Ser Pro Gln Pro Gly Val Pro Ala Asp
        35                  40                  45
Glu Ala Ala Ala Ala Ile Ala Ala Glu Ser Ser Thr Gly Thr Trp Thr
    50                  55                  60
Thr Val Trp Thr Asp Leu Leu Thr Asp Met Asp Arg Tyr Lys Gly Lys
65                  70                  75                  80
Ala Tyr His Val Glu Pro Val Gln Gly Glu Glu Asn Ser Tyr Phe Ala
                85                  90                  95
Phe Ile Ala Tyr Pro Leu Asp Leu Phe Glu Glu Gly Ser Val Thr Asn
            100                 105                 110
Ile Leu Thr Ser Ile Val Gly Asn Val Phe Gly Phe Lys Ala Ile Arg
        115                 120                 125
Ser Leu Arg Leu Glu Asp Ile Arg Phe Pro Val Ala Leu Val Lys Thr
    130                 135                 140
Phe Gln Gly Pro Pro His Gly Ile Gln Val Glu Arg Asp Leu Leu Asn
145                 150                 155                 160
Lys Tyr Gly Arg Pro Met Leu Gly Cys Thr Ile Lys Pro Lys Leu Gly
                165                 170                 175
Leu Ser Ala Lys Asn Tyr Gly Arg Ala Val Tyr Glu Cys Leu Arg Gly
            180                 185                 190
Gly Leu Asp Phe Thr Lys Asp Asp Glu Asn Ile Asn Ser Gln Pro Phe
        195                 200                 205
Gln Arg Trp Arg Asp Arg Phe Leu Phe Val Ala Asp Ala Ile His Lys
    210                 215                 220
Ser Gln Ala Glu Thr Gly Glu Ile Lys Gly His Tyr Leu Asn Val Thr
```

```
225                 230                 235                 240
Ala Pro Thr Cys Glu Glu Met Met Lys Arg Ala Glu Phe Ala Lys Glu
                245                 250                 255
Leu Gly Met Pro Ile Ile Met His Asp Phe Leu Thr Ala Gly Phe Thr
            260                 265                 270
Ala Asn Thr Thr Leu Ala Lys Trp Cys Arg Asp Asn Gly Val Leu Leu
        275                 280                 285
His Ile His Arg Ala Met His Ala Val Ile Asp Arg Gln Arg Asn His
    290                 295                 300
Gly Ile His Phe Arg Val Leu Ala Lys Cys Leu Arg Leu Ser Gly Gly
305                 310                 315                 320
Asp His Leu His Ser Gly Thr Val Val Gly Lys Leu Glu Gly Asp Lys
                325                 330                 335
Ala Ser Thr Leu Gly Phe Val Asp Leu Met Arg Glu Asp His Ile Glu
            340                 345                 350
Ala Asp Arg Ser Arg Gly Val Phe Phe Thr Gln Asp Trp Ala Ser Met
        355                 360                 365
Pro Gly Val Leu Pro Val Ala Ser Gly Gly Ile His Val Trp His Met
    370                 375                 380
Pro Ala Leu Val Glu Ile Phe Gly Asp Asp Ser Val Leu Gln Phe Gly
385                 390                 395                 400
Gly Gly Thr Leu Gly His Pro Trp Gly Asn Ala Pro Gly Ala Thr Ala
                405                 410                 415
Asn Arg Val Ala Leu Glu Ala Cys Val Gln Ala Arg Asn Glu Gly Arg
            420                 425                 430
Asp Leu Tyr Arg Glu Gly Gly Asp Ile Leu Arg Glu Ala Gly Lys Trp
        435                 440                 445
Ser Pro Glu Leu Ala Ala Ala Leu Asp Leu Trp Lys Glu Ile Lys Phe
    450                 455                 460
Glu Phe Glu Thr Met Asp Lys Leu
465                 470
```

<210> 33
<211> 1845
<212> DNA
<213> Artificial

<220>
<223> Com base no nucleotídeo PCC6301 rbcLS de Synechococcus Sp. do tipo selvagem

<400> 33

```
atggccaaga cgcaatctgc cgcaggctat aaggccgggg tgaaggacta caaactcacc     60
tattacaccc ccgattacac ccccaaagac actgacctgc tggcggcttt ccgcttcagc    120
cctcagccgg gtgtccctgc tgacgaagct agcgcggcga tcgcggctga atcttcgacc    180
ggtacctgga ccaccgtgtg gaccgacttg ctgaccgaca tggatcggta caaaggcaag    240
gcctaccacg tggagccggt gcaaggcgaa gagaactcct actttgcgtt catcgcttac    300
ccgctcgacc tgtttgaaga agggtcggtc accaacatcc tgacctcgat cgtcggtaac    360
gtgtttggct tcaaagctat ccgttcgctg cgtctggaag acatccgctt ccccgtcgcc    420
ttggtcaaaa ccttccaagg tcctccccac ggtatccaag tcgagcgcga cctgctgaac    480
aagtacggcc gtccgatgct gggttgcacg atcaaaccaa aactcggtct gtcggcgaaa    540
aactacggtc gtgccgtcta cgaatgtctg cgcggcggtc tggacttcac caaagacgac    600
gaaaacatca actcgcagcc gttccaacgc tggcgcgatc gcttcctgtt tgtggctgat    660
gcaatccaca aatcgcaagc agaaaccggt gaaatcaaag gtcactacct gaacgtgacc    720
gcgccgacct gcgaagaaat gatgaaacgg gctgagttcg ctaaagaact cggcatgccg    780
atcatcatgc atgacttctt gacggctggt ttcaccgcca acaccacctt ggcaaaatgg    840
tgccgcgaca acggcgtcct gctgcacatc caccgtgcaa tgcacgcggt gatcgaccgt    900
cagcgtaacc acgggattca cttccgtgtc ttggccaagt gtttgcgtct gtccggtggt    960
gaccacctcc actccggcac cgtcgtcggc aaactggaag gcgacaaagc ttcgaccttg   1020
ggctttgttg acttgatgcg cgaagaccac atcgaagctg accgcagccg tggggtcttc   1080
ttcacccaag attgggcgtc gatgccgggc gtgctgccgg ttgcttccgg tggtatccac   1140
```

```
gtgtggcaca tgcccgcact ggtggaaatc ttcggtgatg actccgttct ccagttcggt   1200
ggcggcacct tgggtcaccc ctggggtaat gctcctggtg caaccgcgaa ccgtgttgcc   1260
ttggaagctt gcgtccaagc tcggaacgaa ggtcgcgacc tctaccgtga aggcggcgac   1320
atccttcgtg aagctggcaa gtggtcgcct gaactggctg ctgccctcga cctctggaaa   1380
gagatcaagt tcgaattcga aacgatggac aagctctaag gagcctctga ctatcgctgg   1440
gggagtgagc gttgctgcgt aaagctttct ccccagcctt tcgacttaac ctttcaggat   1500
ttctgaatca tgagcatgaa aactctgccc aaagagcgtc gtttcgagac tttctcgtac   1560
ctgcctcccc tcagcgatcg ccaaatcgct gcacaaatcg agtacatgat cgagcaaggc   1620
ttccacccct tgatcgagtt caacgagcac tcgaatccgg aagagttcta ctggacgatg   1680
tggaagctcc ccctgtttga ctgcaagagc cctcagcaag tcctcgatga agtgcgtgag   1740
tgccgcagcg aatacggtga ttgctacatc cgtgtcgctg gcttcgacaa catcaagcag   1800
tgccaaaccg tgagcttcat cgttcatcgt cccggccgat actaa                   1845
```

<210> 34
<211> 472
<212> PRT
<213> Artificial

<220>
<223> Com base na sequência traduzida de PCC6301 rbcLS de Synechococcus Sp. do tipo selvagem

<400> 34

```
Met Ala Lys Thr Gln Ser Ala Ala Gly Tyr Lys Ala Gly Val Lys Asp
1               5                  10                  15
Tyr Lys Leu Thr Tyr Tyr Thr Pro Asp Tyr Thr Pro Lys Asp Thr Asp
            20                  25                  30
Leu Leu Ala Ala Phe Arg Phe Ser Pro Gln Pro Gly Val Pro Ala Asp
        35                  40                  45
Glu Ala Ser Ala Ala Ile Ala Ala Glu Ser Ser Thr Gly Thr Trp Thr
    50                  55                  60
Thr Val Trp Thr Asp Leu Leu Thr Asp Met Asp Arg Tyr Lys Gly Lys
65                  70                  75                  80
Ala Tyr His Val Glu Pro Val Gln Gly Glu Glu Asn Ser Tyr Phe Ala
                85                  90                  95
Phe Ile Ala Tyr Pro Leu Asp Leu Phe Glu Glu Gly Ser Val Thr Asn
            100                 105                 110
Ile Leu Thr Ser Ile Val Gly Asn Val Phe Gly Phe Lys Ala Ile Arg
        115                 120                 125
Ser Leu Arg Leu Glu Asp Ile Arg Phe Pro Val Ala Leu Val Lys Thr
    130                 135                 140
Phe Gln Gly Pro Pro His Gly Ile Gln Val Glu Arg Asp Leu Leu Asn
145                 150                 155                 160
Lys Tyr Gly Arg Pro Met Leu Gly Cys Thr Ile Lys Pro Lys Leu Gly
                165                 170                 175
Leu Ser Ala Lys Asn Tyr Gly Arg Ala Val Tyr Glu Cys Leu Arg Gly
            180                 185                 190
Gly Leu Asp Phe Thr Lys Asp Asp Glu Asn Ile Asn Ser Gln Pro Phe
        195                 200                 205
Gln Arg Trp Arg Asp Arg Phe Leu Phe Val Ala Asp Ala Ile His Lys
    210                 215                 220
Ser Gln Ala Glu Thr Gly Glu Ile Lys Gly His Tyr Leu Asn Val Thr
225                 230                 235                 240
Ala Pro Thr Cys Glu Glu Met Met Lys Arg Ala Glu Phe Ala Lys Glu
                245                 250                 255
Leu Gly Met Pro Ile Ile Met His Asp Phe Leu Thr Ala Gly Phe Thr
            260                 265                 270
Ala Asn Thr Thr Leu Ala Lys Trp Cys Arg Asp Asn Gly Val Leu Leu
        275                 280                 285
His Ile His Arg Ala Met His Ala Val Ile Asp Arg Gln Arg Asn His
```

```
    290                 295                 300
Gly Ile His Phe Arg Val Leu Ala Lys Cys Leu Arg Leu Ser Gly Gly
305                 310                 315                 320
Asp His Leu His Ser Gly Thr Val Val Gly Lys Leu Glu Gly Asp Lys
                325                 330                 335
Ala Ser Thr Leu Gly Phe Val Asp Leu Met Arg Glu Asp His Ile Glu
            340                 345                 350
Ala Asp Arg Ser Arg Gly Val Phe Phe Thr Gln Asp Trp Ala Ser Met
        355                 360                 365
Pro Gly Val Leu Pro Val Ala Ser Gly Gly Ile His Val Trp His Met
    370                 375                 380
Pro Ala Leu Val Glu Ile Phe Gly Asp Asp Ser Val Leu Gln Phe Gly
385                 390                 395                 400
Gly Gly Thr Leu Gly His Pro Trp Gly Asn Ala Pro Gly Ala Thr Ala
                405                 410                 415
Asn Arg Val Ala Leu Glu Ala Cys Val Gln Ala Arg Asn Glu Gly Arg
            420                 425                 430
Asp Leu Tyr Arg Glu Gly Gly Asp Ile Leu Arg Glu Ala Gly Lys Trp
        435                 440                 445
Ser Pro Glu Leu Ala Ala Ala Leu Asp Leu Trp Lys Glu Ile Lys Phe
    450                 455                 460
Glu Phe Glu Thr Met Asp Lys Leu
465                 470
```

<210> 35
<211> 1845
<212> DNA
<213> Artificial

<220>
<223> Com base no nucleotídeo PCC6301 rbcLS de Synechococcus Sp. do tipo selvagem

<400> 35

```
atggccaaga cgcaatctgc cgcaggctat aaggccgggg tgaaggacta caaactcggc     60
tattggaccc ccgattacac ccccaaagac actgacctgc tggcggcttt ccgcttcagc    120
cctcagccgg gtgtccctgc tatcgaagct ggtgcggcga tcgcggctga atcttcgacc    180
ggtacctgga ccaccgtgtg gaccgacttg ctgaccgaca tggatcggta ccgcggcaag    240
tgctaccaca tcgagccggt gcaaggcgaa gagaactcct actttgcgtt catcgcttac    300
ccgctcgacc tgtttgaaga agggtcggtc accaacatcc tgacctcgat cgtcggtaac    360
gtgtttggct tcaaagctat ccgttcgctg cgtctggaag acatccgctt ccccgtcgcc    420
ttggtcaaaa ccttccaagg tcctccccac ggtatccaag tcgagcgcga cctgctgaac    480
aagtacggcc gtccgatgct gggttgcacg atcaaaccaa aactcggtct gtcggcgaaa    540
aactacggtc gtgccgtcta cgaatgtctg cgcggcggtc tggacttcac caaagacgac    600
gaaaacatca actcgcagcc gttccaacgc tggcgcgatc gcttcctgtt tgtggctgat    660
gcaatccaca aatcgcaagc agaaaccggt gaaatcaaag gtcactacct gaacgtgacc    720
gcgccgacct gcgaagaaat gatgaaacgg gctgagttcg ctaaagaact cggcatgccg    780
atcatcatgc atgacttctt gacggctggt ttcaccgcca acaccacctt ggcaaaatgg    840
tgccgcgaca acggcgtcct gctgcacatc caccgtgcaa tgcacgcggt gatcgaccgt    900
cagcgtaacc acgggattca cttccgtgtc ttggccaagt gtttgcgtct gtccggtggt    960
gaccacctcc actccggcac cgtcgtcggc aaactggaag gcgacaaagc ttcgaccttg   1020
ggctttgttg acttgatgcg cgaagaccac atcgaagctg accgcagccg tggggtcttc   1080
ttcacccaag attgggcgtc gatgccgggc gtgctgccgg ttgcttccgg tggtatccac   1140
gtgtggcaca tgcccgcact ggtggaaatc ttcggtgatg actccgttct ccagttcggt   1200
ggcggcacct tgggtcaccc ctggggtaat gctcctggtg caaccgcgaa ccgtgttgcc   1260
ttggaagctt gcgtccaagc tcggaacgaa ggtcgcgacc tctaccgtga aggcggcgac   1320
atccttcgtg aagctggcaa gtggtcgcct gaactggctg ctgccctcga cctctggaaa   1380
gagatcaagt tcgaattcga aacgatggac aagctctaag gagcctctga ctatcgctgg   1440
gggagtgagc gttgctgcgt aaagctttct ccccagcctt tcgacttaac ctttcaggat   1500
ttctgaatca tgagcatgaa aactctgccc aaagagcgtc gtttcgagac tttctcgtac   1560
ctgcctcccc tcagcgatcg ccaaatcgct gcacaaatcg agtacatgat cgagcaaggc   1620
```

```
ttccacccct tgatcgagtt caacgagcac tcgaatccgg aagagttcta ctggacgatg   1680
tggaagctcc ccctgtttga ctgcaagagc cctcagcaag tcctcgatga agtgcgtgag   1740
tgccgcagcg aatacggtga ttgctacatc cgtgtcgctg gcttcgacaa catcaagcag   1800
tgccaaaccg tgagcttcat cgttcatcgt cccggccgat actaa                   1845
```

<210> 36
<211> 472
<212> PRT
<213> Artificial

<220>
<223> Com base na sequência traduzida de PCC6301 rbcLS de Synechococcus Sp. do tipo selvagem

<400> 36

```
Met Ala Lys Thr Gln Ser Ala Ala Gly Tyr Lys Ala Gly Val Lys Asp
1               5                  10                  15
Tyr Lys Leu Gly Tyr Trp Thr Pro Asp Tyr Thr Pro Lys Asp Thr Asp
            20                  25                  30
Leu Leu Ala Ala Phe Arg Phe Ser Pro Gln Pro Gly Val Pro Ala Ile
        35                  40                  45
Glu Ala Gly Ala Ala Ile Ala Ala Glu Ser Ser Thr Gly Thr Trp Thr
    50                  55                  60
Thr Val Trp Thr Asp Leu Leu Thr Asp Met Asp Arg Tyr Arg Gly Lys
65                  70                  75                  80
Cys Tyr His Ile Glu Pro Val Gln Gly Glu Glu Asn Ser Tyr Phe Ala
                85                  90                  95
Phe Ile Ala Tyr Pro Leu Asp Leu Phe Glu Glu Gly Ser Val Thr Asn
            100                 105                 110
Ile Leu Thr Ser Ile Val Gly Asn Val Phe Gly Phe Lys Ala Ile Arg
        115                 120                 125
Ser Leu Arg Leu Glu Asp Ile Arg Phe Pro Val Ala Leu Val Lys Thr
    130                 135                 140
Phe Gln Gly Pro Pro His Gly Ile Gln Val Glu Arg Asp Leu Leu Asn
145                 150                 155                 160
Lys Tyr Gly Arg Pro Met Leu Gly Cys Thr Ile Lys Pro Lys Leu Gly
                165                 170                 175
Leu Ser Ala Lys Asn Tyr Gly Arg Ala Val Tyr Glu Cys Leu Arg Gly
            180                 185                 190
Gly Leu Asp Phe Thr Lys Asp Asp Glu Asn Ile Asn Ser Gln Pro Phe
        195                 200                 205
Gln Arg Trp Arg Asp Arg Phe Leu Phe Val Ala Asp Ala Ile His Lys
    210                 215                 220
Ser Gln Ala Glu Thr Gly Glu Ile Lys Gly His Tyr Leu Asn Val Thr
225                 230                 235                 240
Ala Pro Thr Cys Glu Glu Met Met Lys Arg Ala Glu Phe Ala Lys Glu
                245                 250                 255
Leu Gly Met Pro Ile Ile Met His Asp Phe Leu Thr Ala Gly Phe Thr
            260                 265                 270
Ala Asn Thr Thr Leu Ala Lys Trp Cys Arg Asp Asn Gly Val Leu Leu
        275                 280                 285
His Ile His Arg Ala Met His Ala Val Ile Asp Arg Gln Arg Asn His
    290                 295                 300
Gly Ile His Phe Arg Val Leu Ala Lys Cys Leu Arg Leu Ser Gly Gly
305                 310                 315                 320
Asp His Leu His Ser Gly Thr Val Val Gly Lys Leu Glu Gly Asp Lys
                325                 330                 335
Ala Ser Thr Leu Gly Phe Val Asp Leu Met Arg Glu Asp His Ile Glu
            340                 345                 350
Ala Asp Arg Ser Arg Gly Val Phe Phe Thr Gln Asp Trp Ala Ser Met
```

```
        355                 360                 365
Pro Gly Val Leu Pro Val Ala Ser Gly Gly Ile His Val Trp His Met
    370                 375                 380
Pro Ala Leu Val Glu Ile Phe Gly Asp Asp Ser Val Leu Gln Phe Gly
385                 390                 395                 400
Gly Gly Thr Leu Gly His Pro Trp Gly Asn Ala Pro Gly Ala Thr Ala
                405                 410                 415
Asn Arg Val Ala Leu Glu Ala Cys Val Gln Ala Arg Asn Glu Gly Arg
            420                 425                 430
Asp Leu Tyr Arg Glu Gly Gly Asp Ile Leu Arg Glu Ala Gly Lys Trp
        435                 440                 445
Ser Pro Glu Leu Ala Ala Ala Leu Asp Leu Trp Lys Glu Ile Lys Phe
    450                 455                 460
Glu Phe Glu Thr Met Asp Lys Leu
465                 470
```

<210> 37
<211> 1845
<212> DNA
<213> Artificial

<220>
<223> Com base no nucleotídeo PCC6301 rbcLS de Synechococcus Sp. do tipo selvagem

<400> 37

```
atggccaaga cgcaatctgc cgcaggctat aaggccgggg tgaaggacta caaactcacc     60
tattggaccc ccgattacac ccccaaagac actgacctgc tggcggcttt ccgcttcagc    120
cctcagccgg gtgtccctgc tatcgaagct gcggcggcga tcgcggctga atcttcgacc    180
ggtacctgga ccaccgtgtg gaccgacttg ctgaccgaca tggatcggta caaaggcaag    240
gcctaccacg tggagccggt gcaaggcgaa gagaactcct actttgcgtt catcgcttac    300
ccgctcgacc tgtttgaaga agggtcggtc accaacatcc tgacctcgat cgtcggtaac    360
gtgtttggct tcaaagctat ccgttcgctg cgtctggaag acatccgctt ccccgtcgcc    420
ttggtcaaaa ccttccaagg tcctccccac ggtatccaag tcgagcgcga cctgctgaac    480
aagtacggcc gtccgatgct gggttgcacg atcaaaccaa aactcggtct gtcggcgaaa    540
aactacggtc gtgccgtcta cgaatgtctg cgcggcggtc tggacttcac caaagacgac    600
gaaaacatca actcgcagcc gttccaacgc tggcgcgatc gcttcctgtt tgtggctgat    660
gcaatccaca aatcgcaagc agaaaccggt gaaatcaaag gtcactacct gaacgtgacc    720
gcgccgacct gcgaagaaat gatgaaacgg gctgagttcg ctaaagaact cggcatgccg    780
atcatcatgc atgacttctt gacggctggt ttcaccgcca acaccacctt ggcaaaatgg    840
tgccgcgaca acggcgtcct gctgcacatc caccgtgcaa tgcacgcggt gatcgaccgt    900
cagcgtaacc acgggattca cttccgtgtc ttggccaagt gtttgcgtct gtccggtggt    960
gaccacctcc actccggcac cgtcgtcggc aaactggaag gcgacaaagc ttcgaccttg   1020
ggctttgttg acttgatgcg cgaagaccac atcgaagctg accgcagccg tggggtcttc   1080
ttcacccaag attgggcgtc gatgccgggc gtgctgccgg ttgcttccgg tggtatccac   1140
gtgtggcaca tgcccgcact ggtggaaatc ttcggtgatg actccgttct ccagttcggt   1200
ggcggcacct tgggtcaccc ctggggtaat gctcctggtg caaccgcgaa ccgtgttgcc   1260
ttggaagctt gcgtccaagc tcggaacgaa ggtcgcgacc tctaccgtga aggcggcgac   1320
atccttcgtg aagctggcaa gtggtcgcct gaactggctg ctgccctcga cctctggaaa   1380
gagatcaagt tcgaattcga aacgatggac aagctctaag gagcctctga ctatcgctgg   1440
gggagtgagc gttgctgcgt aaagctttct ccccagcctt tcgacttaac ctttcaggat   1500
ttctgaatca tgagcatgaa aactctgccc aaagagcgtc gtttcgagac tttctcgtac   1560
ctgcctcccc tcagcgatcg ccaaatcgct gcacaaatcg agtacatgat cgagcaaggc   1620
ttccaccccc tgatcgagtt caacgagcac tcgaatccgg aagagttcta ctggacgatg   1680
tggaagctcc ccctgtttga ctgcaagagc cctcagcaag tcctcgatga agtgcgtgag   1740
tgccgcagcg aatacggtga ttgctacatc cgtgtcgctg gcttcgacaa catcaagcag   1800
tgccaaaccg tgagcttcat cgttcatcgt cccggccgat actaa                   1845
```

<210> 38
<211> 472
<212> PRT

<213> Artificial

<220>
<223> Com base na sequência traduzida de PCC6301 rbcLS de Synechococcus Sp. do tipo selvagem

<400> 38

```
Met Ala Lys Thr Gln Ser Ala Ala Gly Tyr Lys Ala Gly Val Lys Asp
1               5                  10                  15
Tyr Lys Leu Thr Tyr Trp Thr Pro Asp Tyr Thr Pro Lys Asp Thr Asp
            20                  25                  30
Leu Leu Ala Ala Phe Arg Phe Ser Pro Gln Pro Gly Val Pro Ala Ile
        35                  40                  45
Glu Ala Ala Ala Ala Ile Ala Ala Glu Ser Ser Thr Gly Thr Trp Thr
    50                  55                  60
Thr Val Trp Thr Asp Leu Leu Thr Asp Met Asp Arg Tyr Lys Gly Lys
65                  70                  75                  80
Ala Tyr His Val Glu Pro Val Gln Gly Glu Glu Asn Ser Tyr Phe Ala
                85                  90                  95
Phe Ile Ala Tyr Pro Leu Asp Leu Phe Glu Glu Gly Ser Val Thr Asn
            100                 105                 110
Ile Leu Thr Ser Ile Val Gly Asn Val Phe Gly Phe Lys Ala Ile Arg
        115                 120                 125
Ser Leu Arg Leu Glu Asp Ile Arg Phe Pro Val Ala Leu Val Lys Thr
    130                 135                 140
Phe Gln Gly Pro Pro His Gly Ile Gln Val Glu Arg Asp Leu Leu Asn
145                 150                 155                 160
Lys Tyr Gly Arg Pro Met Leu Gly Cys Thr Ile Lys Pro Lys Leu Gly
                165                 170                 175
Leu Ser Ala Lys Asn Tyr Gly Arg Ala Val Tyr Glu Cys Leu Arg Gly
            180                 185                 190
Gly Leu Asp Phe Thr Lys Asp Asp Glu Asn Ile Asn Ser Gln Pro Phe
        195                 200                 205
Gln Arg Trp Arg Asp Arg Phe Leu Phe Val Ala Asp Ala Ile His Lys
    210                 215                 220
Ser Gln Ala Glu Thr Gly Glu Ile Lys Gly His Tyr Leu Asn Val Thr
225                 230                 235                 240
Ala Pro Thr Cys Glu Glu Met Met Lys Arg Ala Glu Phe Ala Lys Glu
                245                 250                 255
Leu Gly Met Pro Ile Ile Met His Asp Phe Leu Thr Ala Gly Phe Thr
            260                 265                 270
Ala Asn Thr Thr Leu Ala Lys Trp Cys Arg Asp Asn Gly Val Leu Leu
        275                 280                 285
His Ile His Arg Ala Met His Ala Val Ile Asp Arg Gln Arg Asn His
    290                 295                 300
Gly Ile His Phe Arg Val Leu Ala Lys Cys Leu Arg Leu Ser Gly Gly
305                 310                 315                 320
Asp His Leu His Ser Gly Thr Val Val Gly Lys Leu Glu Gly Asp Lys
                325                 330                 335
Ala Ser Thr Leu Gly Phe Val Asp Leu Met Arg Glu Asp His Ile Glu
            340                 345                 350
Ala Asp Arg Ser Arg Gly Val Phe Phe Thr Gln Asp Trp Ala Ser Met
        355                 360                 365
Pro Gly Val Leu Pro Val Ala Ser Gly Gly Ile His Val Trp His Met
    370                 375                 380
Pro Ala Leu Val Glu Ile Phe Gly Asp Asp Ser Val Leu Gln Phe Gly
385                 390                 395                 400
Gly Gly Thr Leu Gly His Pro Trp Gly Asn Ala Pro Gly Ala Thr Ala
                405                 410                 415
Asn Arg Val Ala Leu Glu Ala Cys Val Gln Ala Arg Asn Glu Gly Arg
```

```
            420                 425                 430
Asp Leu Tyr Arg Glu Gly Gly Asp Ile Leu Arg Glu Ala Gly Lys Trp
        435                 440                 445
Ser Pro Glu Leu Ala Ala Ala Leu Asp Leu Trp Lys Glu Ile Lys Phe
    450                 455                 460
Glu Phe Glu Thr Met Asp Lys Leu
465                 470
```

<210> 39
<211> 1845
<212> DNA
<213> Artificial

<220>
<223> Com base no nucleotídeo PCC6301 rbcLS de Synechococcus Sp. do tipo selvagem

<400> 39

```
atggccaaga cgcaatctgc cgcaggctat aaggccgggg tgaaggacta caaactcacc     60
tattggaccc ccgattacac ccccaaagac actgacctgc tggcggcttt ccgcttcagc    120
cctcagccgg gtgtccctgc tatcgaagct agcgcggcga tcgcggctga atcttcgacc    180
ggtacctgga ccaccgtgtg gaccgacttg ctgaccgaca tggatcggta caaaggcaag    240
gcctaccacg tggagccggt gcaaggcgaa gagaactcct actttgcgtt catcgcttac    300
ccgctcgacc tgtttgaaga agggtcggtc accaacatcc tgacctcgat cgtcggtaac    360
gtgtttggct tcaaagctat ccgttcgctg cgtctggaag acatccgctt ccccgtcgcc    420
ttggtcaaaa ccttccaagg tcctccccac ggtatccaag tcgagcgcga cctgctgaac    480
aagtacggcc gtccgatgct gggttgcacg atcaaaccaa aactcggtct gtcggcgaaa    540
aactacggtc gtgccgtcta cgaatgtctg cgcggcggtc tggacttcac caaagacgac    600
gaaaacatca actcgcagcc gttccaacgc tggcgcgatc gcttcctgtt tgtggctgat    660
gcaatccaca aatcgcaagc agaaaccggt gaaatcaaag gtcactacct gaacgtgacc    720
gcgccgacct gcgaagaaat gatgaaacgg gctgagttcg ctaaagaact cggcatgccg    780
atcatcatgc atgacttctt gacggctggt ttcaccgcca acaccacctt ggcaaaatgg    840
tgccgcgaca acggcgtcct gctgcacatc caccgtgcaa tgcacgcggt gatcgaccgt    900
cagcgtaacc acgggattca cttccgtgtc ttggccaagt gtttgcgtct gtccggtggt    960
gaccacctcc actccggcac cgtcgtcggc aaactggaag gcgacaaagc ttcgaccttg   1020
ggctttgttg acttgatgcg cgaagaccac atcgaagctg accgcagccg tggggtcttc   1080
ttcacccaag attgggcgtc gatgccgggc gtgctgccgg ttgcttccgg tggtatccac   1140
gtgtggcaca tgcccgcact ggtggaaatc ttcggtgatg actccgttct ccagttcggt   1200
ggcggcacct tgggtcaccc ctggggtaat gctcctggtg caaccgcgaa ccgtgttgcc   1260
ttggaagctt gcgtccaagc tcggaacgaa ggtcgcgacc tctaccgtga aggcggcgac   1320
atccttcgtg aagctggcaa gtggtcgcct gaactggctg ctgccctcga cctctggaaa   1380
gagatcaagt tcgaattcga aacgatggac aagctctaag gagcctctga ctatcgctgg   1440
gggagtgagc gttgctgcgt aaagctttct ccccagcctt tcgacttaac ctttcaggat   1500
ttctgaatca tgagcatgaa aactctgccc aaagagcgtc gtttcgagac tttctcgtac   1560
ctgcctcccc tcagcgatcg ccaaatcgct gcacaaatcg agtacatgat cgagcaaggc   1620
ttccaccect tgatcgagtt caacgagcac tcgaatccgg aagagttcta ctggacgatg   1680
tggaagctcc ccctgtttga ctgcaagagc cctcagcaag tcctcgatga agtgcgtgag   1740
tgccgcagcg aatacggtga ttgctacatc cgtgtcgctg gcttcgacaa catcaagcag   1800
tgccaaaccg tgagcttcat cgttcatcgt cccggccgat actaa                   1845
```

<210> 40
<211> 472
<212> PRT
<213> Artificial

<220>
<223> Com base na sequência traduzida de PCC6301 rbcLS de Synechococcus Sp. do tipo selvagem

<400> 40

```
Met Ala Lys Thr Gln Ser Ala Ala Gly Tyr Lys Ala Gly Val Lys Asp
1               5                  10                  15
Tyr Lys Leu Thr Tyr Trp Thr Pro Asp Tyr Thr Pro Lys Asp Thr Asp
            20                  25                  30
Leu Leu Ala Ala Phe Arg Phe Ser Pro Gln Pro Gly Val Pro Ala Ile
        35                  40                  45
Glu Ala Ser Ala Ala Ile Ala Ala Glu Ser Ser Thr Gly Thr Trp Thr
    50                  55                  60
Thr Val Trp Thr Asp Leu Leu Thr Asp Met Asp Arg Tyr Lys Gly Lys
65                  70                  75                  80
Ala Tyr His Val Glu Pro Val Gln Gly Glu Glu Asn Ser Tyr Phe Ala
                85                  90                  95
Phe Ile Ala Tyr Pro Leu Asp Leu Phe Glu Glu Gly Ser Val Thr Asn
            100                 105                 110
Ile Leu Thr Ser Ile Val Gly Asn Val Phe Gly Phe Lys Ala Ile Arg
        115                 120                 125
Ser Leu Arg Leu Glu Asp Ile Arg Phe Pro Val Ala Leu Val Lys Thr
    130                 135                 140
Phe Gln Gly Pro Pro His Gly Ile Gln Val Glu Arg Asp Leu Leu Asn
145                 150                 155                 160
Lys Tyr Gly Arg Pro Met Leu Gly Cys Thr Ile Lys Pro Lys Leu Gly
                165                 170                 175
Leu Ser Ala Lys Asn Tyr Gly Arg Ala Val Tyr Glu Cys Leu Arg Gly
            180                 185                 190
Gly Leu Asp Phe Thr Lys Asp Asp Glu Asn Ile Asn Ser Gln Pro Phe
        195                 200                 205
Gln Arg Trp Arg Asp Arg Phe Leu Phe Val Ala Asp Ala Ile His Lys
    210                 215                 220
Ser Gln Ala Glu Thr Gly Glu Ile Lys Gly His Tyr Leu Asn Val Thr
225                 230                 235                 240
Ala Pro Thr Cys Glu Glu Met Met Lys Arg Ala Glu Phe Ala Lys Glu
                245                 250                 255
Leu Gly Met Pro Ile Ile Met His Asp Phe Leu Thr Ala Gly Phe Thr
            260                 265                 270
Ala Asn Thr Thr Leu Ala Lys Trp Cys Arg Asp Asn Gly Val Leu Leu
        275                 280                 285
His Ile His Arg Ala Met His Ala Val Ile Asp Arg Gln Arg Asn His
    290                 295                 300
Gly Ile His Phe Arg Val Leu Ala Lys Cys Leu Arg Leu Ser Gly Gly
305                 310                 315                 320
Asp His Leu His Ser Gly Thr Val Val Gly Lys Leu Glu Gly Asp Lys
                325                 330                 335
Ala Ser Thr Leu Gly Phe Val Asp Leu Met Arg Glu Asp His Ile Glu
            340                 345                 350
Ala Asp Arg Ser Arg Gly Val Phe Phe Thr Gln Asp Trp Ala Ser Met
        355                 360                 365
Pro Gly Val Leu Pro Val Ala Ser Gly Gly Ile His Val Trp His Met
    370                 375                 380
Pro Ala Leu Val Glu Ile Phe Gly Asp Asp Ser Val Leu Gln Phe Gly
385                 390                 395                 400
Gly Gly Thr Leu Gly His Pro Trp Gly Asn Ala Pro Gly Ala Thr Ala
                405                 410                 415
Asn Arg Val Ala Leu Glu Ala Cys Val Gln Ala Arg Asn Glu Gly Arg
            420                 425                 430
Asp Leu Tyr Arg Glu Gly Gly Asp Ile Leu Arg Glu Ala Gly Lys Trp
        435                 440                 445
Ser Pro Glu Leu Ala Ala Ala Leu Asp Leu Trp Lys Glu Ile Lys Phe
    450                 455                 460
Glu Phe Glu Thr Met Asp Lys Leu
465                 470
```

<210> 41
<211> 1845
<212> DNA
<213> Artificial

<220>
<223> Com base no nucleotídeo PCC6301 rbcLS de Synechococcus Sp. do tipo selvagem

<400> 41
atggccaaga cgcaatctgc cgcaggctat aaggccgggg tgaaggacta caaactcacc 60
tattggaccc ccgattacac ccccaaagac actgacctgc tggcggcttt ccgcttcagc 120
cctcagccgg gtgtccctgc tgtcgaagct gcggcggcga tcgcggctga atcttcgacc 180
ggtacctgga ccaccgtgtg gaccgacttg ctgaccgaca tggatcggta caaaggcaag 240
gcctaccacg tggagccggt gcaaggcgaa gagaactcct actttgcgtt catcgcttac 300
ccgctcgacc tgtttgaaga agggtcggtc accaacatcc tgacctcgat cgtcggtaac 360
gtgtttggct tcaaagctat ccgttcgctg cgtctggaag acatccgctt ccccgtcgcc 420
ttggtcaaaa ccttccaagg tcctccccac ggtatccaag tcgagcgcga cctgctgaac 480
aagtacggcc gtccgatgct gggttgcacg atcaaaccaa aactcggtct gtcggcgaaa 540
aactacggtc gtgccgtcta cgaatgtctg cgcggcggtc tggacttcac caaagacgac 600
gaaaacatca actcgcagcc gttccaacgc tggcgcgatc gcttcctgtt tgtggctgat 660
gcaatccaca aatcgcaagc agaaaccggt gaaatcaaag gtcactacct gaacgtgacc 720
gcgccgacct gcgaagaaat gatgaaacgg gctgagttcg ctaaagaact cggcatgccg 780
atcatcatgc atgacttctt gacggctggt ttcaccgcca acaccacctt ggcaaaatgg 840
tgccgcgaca acggcgtcct gctgcacatc caccgtgcaa tgcacgcggt gatcgaccgt 900
cagcgtaacc acgggattca cttccgtgtc ttggccaagt gtttgcgtct gtccggtggt 960
gaccacctcc actccggcac cgtcgtcggc aaactggaag gcgacaaagc ttcgaccttg 1020
ggctttgttg acttgatgcg cgaagaccac atcgaagctg accgcagccg tggggtcttc 1080
ttcacccaag attgggcgtc gatgccgggc gtgctgccgg ttgcttccgg tggtatccac 1140
gtgtggcaca tgcccgcact ggtggaaatc ttcggtgatg actccgttct ccagttcggt 1200
ggcggcacct tgggtcaccc ctggggtaat gctcctggtg caaccgcgaa ccgtgttgcc 1260
ttggaagctt gcgtccaagc tcggaacgaa ggtcgcgacc tctaccgtga aggcggcgac 1320
atccttcgtg aagctggcaa gtggtcgcct gaactggctg ctgccctcga cctctggaaa 1380
gagatcaagt tcgaattcga aacgatggac aagctctaag gagcctctga ctatcgctgg 1440
gggagtgagc gttgctgcgt aaagctttct ccccagcctt tcgacttaac ctttcaggat 1500
ttctgaatca tgagcatgaa aactctgccc aaagagcgtc gtttcgagac tttctcgtac 1560
ctgcctcccc tcagcgatcg ccaaatcgct gcacaaatcg agtacatgat cgagcaaggc 1620
ttccaccct tgatcgagtt caacgagcac tcgaatccgg aagagttcta ctggacgatg 1680
tggaagctcc ccctgtttga ctgcaagagc cctcagcaag tcctcgatga agtgcgtgag 1740
tgccgcagcg aatacggtga ttgctacatc cgtgtcgctg gcttcgacaa catcaagcag 1800
tgccaaaccg tgagcttcat cgttcatcgt cccggccgat actaa 1845

<210> 42
<211> 472
<212> PRT
<213> Artificial

<220>
<223> Com base na sequência traduzida de PCC6301 rbcLS de Synechococcus Sp. do tipo selvagem

<400> 42

Met Ala Lys Thr Gln Ser Ala Ala Gly Tyr Lys Ala Gly Val Lys Asp
1 5 10 15
Tyr Lys Leu Thr Tyr Trp Thr Pro Asp Tyr Thr Pro Lys Asp Thr Asp
20 25 30
Leu Leu Ala Ala Phe Arg Phe Ser Pro Gln Pro Gly Val Pro Ala Val
35 40 45
Glu Ala Ala Ala Ala Ile Ala Ala Glu Ser Ser Thr Gly Thr Trp Thr
50 55 60

```
Thr Val Trp Thr Asp Leu Leu Thr Asp Met Asp Arg Tyr Lys Gly Lys
65                  70                  75                  80
Ala Tyr His Val Glu Pro Val Gln Gly Glu Glu Asn Ser Tyr Phe Ala
                85                  90                  95
Phe Ile Ala Tyr Pro Leu Asp Leu Phe Glu Glu Gly Ser Val Thr Asn
            100                 105                 110
Ile Leu Thr Ser Ile Val Gly Asn Val Phe Gly Phe Lys Ala Ile Arg
        115                 120                 125
Ser Leu Arg Leu Glu Asp Ile Arg Phe Pro Val Ala Leu Val Lys Thr
    130                 135                 140
Phe Gln Gly Pro Pro His Gly Ile Gln Val Glu Arg Asp Leu Leu Asn
145                 150                 155                 160
Lys Tyr Gly Arg Pro Met Leu Gly Cys Thr Ile Lys Pro Lys Leu Gly
                165                 170                 175
Leu Ser Ala Lys Asn Tyr Gly Arg Ala Val Tyr Glu Cys Leu Arg Gly
            180                 185                 190
Gly Leu Asp Phe Thr Lys Asp Asp Glu Asn Ile Asn Ser Gln Pro Phe
        195                 200                 205
Gln Arg Trp Arg Asp Arg Phe Leu Phe Val Ala Asp Ala Ile His Lys
    210                 215                 220
Ser Gln Ala Glu Thr Gly Glu Ile Lys Gly His Tyr Leu Asn Val Thr
225                 230                 235                 240
Ala Pro Thr Cys Glu Glu Met Met Lys Arg Ala Glu Phe Ala Lys Glu
                245                 250                 255
Leu Gly Met Pro Ile Ile Met His Asp Phe Leu Thr Ala Gly Phe Thr
            260                 265                 270
Ala Asn Thr Thr Leu Ala Lys Trp Cys Arg Asp Asn Gly Val Leu Leu
        275                 280                 285
His Ile His Arg Ala Met His Ala Val Ile Asp Arg Gln Arg Asn His
    290                 295                 300
Gly Ile His Phe Arg Val Leu Ala Lys Cys Leu Arg Leu Ser Gly Gly
305                 310                 315                 320
Asp His Leu His Ser Gly Thr Val Val Gly Lys Leu Glu Gly Asp Lys
                325                 330                 335
Ala Ser Thr Leu Gly Phe Val Asp Leu Met Arg Glu Asp His Ile Glu
            340                 345                 350
Ala Asp Arg Ser Arg Gly Val Phe Phe Thr Gln Asp Trp Ala Ser Met
        355                 360                 365
Pro Gly Val Leu Pro Val Ala Ser Gly Gly Ile His Val Trp His Met
    370                 375                 380
Pro Ala Leu Val Glu Ile Phe Gly Asp Asp Ser Val Leu Gln Phe Gly
385                 390                 395                 400
Gly Gly Thr Leu Gly His Pro Trp Gly Asn Ala Pro Gly Ala Thr Ala
                405                 410                 415
Asn Arg Val Ala Leu Glu Ala Cys Val Gln Ala Arg Asn Glu Gly Arg
            420                 425                 430
Asp Leu Tyr Arg Glu Gly Gly Asp Ile Leu Arg Glu Ala Gly Lys Trp
        435                 440                 445
Ser Pro Glu Leu Ala Ala Ala Leu Asp Leu Trp Lys Glu Ile Lys Phe
    450                 455                 460
Glu Phe Glu Thr Met Asp Lys Leu
465                 470
```

<210> 43
<211> 1845
<212> DNA
<213> Artificial

<220>
<223> Com base no nucleotídeo PCC6301 rbcLS de Synechococcus Sp. do tipo selvagem

<400> 43

```
atggccaaga cgcaatctgc cgcaggctat aaggccgggg tgaaggacta caaactcacc     60
tattacaccc ccgattacac ccccaaagac actgacctgc tggcggcttt ccgcttcagc    120
cctcagccgg gtgtccctgc tgacgaagct ggtgcggcga tcgcggctga atcttcgacc    180
ggtacctgga ccaccgtgtg gaccgacttg ctgaccgaca tggatcggta caaaggcaag    240
tgctaccaca tcgagccggt gcaaggcgaa gagaactcct actttgcgtt catcgcttac    300
ccgctcgacc tgtttgaaga agggtcggtc accaacatcc tgacctcgat catcggtaat    360
gtgtttggct tcaaagctat ccgttcgctg cgtctggaag acatccgctt ccccgtcgcc    420
ttggtcaaaa ccttccaagg tcctccccac ggtatccaag tcgagcgcga cctgctgaac    480
aagtacggcc gtccgatgct gggttgcacg atcaaaccaa aactcggtct gtcggcgaaa    540
aactacggtc gtgccgtcta cgaatgtctg cgcggcggtc tggacttcac caaagacgac    600
gaaaacatca actcgcagcc gttccaacgc tggcgcgatc gcttcctgtt tgtggctgat    660
gcaatccaca aatcgcaagc agaaaccggt gaaatcaaag gtcactacct gaacgtgacc    720
gcgccgacct gcgaagaaat gatgaaacgg gctgagttcg ctaaagaact cggcatgccg    780
atcatcatgc atgacttctt gacggctggt ttcaccgcca acaccacctt ggcaaaatgg    840
tgccgcgaca acggcgtcct gctgcacatc caccgtgcag ggcacgcgac gatcgaccgt    900
cagcgtaacc acgggattca cttccgtgtc ttggccaagt gtttgcgtct gtccggtggt    960
gaccacctcc actccggcac cgtcgtcggc aaactggaag gcgacaaagc ttcgaccttg   1020
ggctttgttg acttgatgcg cgaagaccac atcgaagctg accgcagccg tggggtcttc   1080
ttcacccaag attgggcgtc gatgccgggc gtgctgccgg ttgcttccgg tggtatccac   1140
gtgtggcaca tgcccgcact ggtggaaatc ttcggtgatg actccgttct ccagttcggt   1200
ggcggcacct tgggtcaccc ctggggtaat gctcctggtg caaccgcgaa ccgtgttgcc   1260
ttggaagctt gcgtccaagc tcggaacgaa ggtcgcgacc tctaccgtga aggcggcgac   1320
atccttcgtg aagctggcaa gtggtcgcct gaactggctg ctgccctcga cctctggaaa   1380
gagatcaagt tcgaattcga aacgatggac aagctctaag gagcctctga ctatcgctgg   1440
gggagtgagc gttgctgcgt aaagctttct ccccagcctt tcgacttaac ctttcaggat   1500
ttctgaatca tgagcatgaa aactctgccc aaagagcgtc gtttcgagac tttctcgtac   1560
ctgcctcccc tcagcgatcg ccaaatcgct gcacaaatcg agtacatgat cgagcaaggc   1620
ttccaccccc tgatcgagtt caacgagcac tcgaatccgg aagagttcta ctggacgatg   1680
tggaagctcc ccctgtttga ctgcaagagc cctcagcaag tcctcgatga agtgcgtgag   1740
tgccgcagcg aatacggtga ttgctacatc cgtgtcgctg gcttcgacaa catcaagcag   1800
tgccaaaccg tgagcttcat cgttcatcgt cccggccgat actaa                   1845
```

<210> 44
<211> 472
<212> PRT
<213> Artificial

<220>
<223> Com base na sequência traduzida de PCC6301 rbcLS de Synechococcus Sp. do tipo selvagem

<400> 44

```
Met Ala Lys Thr Gln Ser Ala Ala Gly Tyr Lys Ala Gly Val Lys Asp
1               5                  10                  15
Tyr Lys Leu Thr Tyr Tyr Thr Pro Asp Tyr Thr Pro Lys Asp Thr Asp
            20                  25                  30
Leu Leu Ala Ala Phe Arg Phe Ser Pro Gln Pro Gly Val Pro Ala Asp
        35                  40                  45
Glu Ala Gly Ala Ala Ile Ala Ala Glu Ser Ser Thr Gly Thr Trp Thr
    50                  55                  60
Thr Val Trp Thr Asp Leu Leu Thr Asp Met Asp Arg Tyr Lys Gly Lys
65                  70                  75                  80
Cys Tyr His Ile Glu Pro Val Gln Gly Glu Glu Asn Ser Tyr Phe Ala
                85                  90                  95
Phe Ile Ala Tyr Pro Leu Asp Leu Phe Glu Glu Gly Ser Val Thr Asn
            100                 105                 110
Ile Leu Thr Ser Ile Ile Gly Asn Val Phe Gly Phe Lys Ala Ile Arg
        115                 120                 125
```

```
Ser Leu Arg Leu Glu Asp Ile Arg Phe Pro Val Ala Leu Val Lys Thr
    130                 135                 140
Phe Gln Gly Pro Pro His Gly Ile Gln Val Glu Arg Asp Leu Leu Asn
145                 150                 155                 160
Lys Tyr Gly Arg Pro Met Leu Gly Cys Thr Ile Lys Pro Lys Leu Gly
                165                 170                 175
Leu Ser Ala Lys Asn Tyr Gly Arg Ala Val Tyr Glu Cys Leu Arg Gly
            180                 185                 190
Gly Leu Asp Phe Thr Lys Asp Asp Glu Asn Ile Asn Ser Gln Pro Phe
        195                 200                 205
Gln Arg Trp Arg Asp Arg Phe Leu Phe Val Ala Asp Ala Ile His Lys
    210                 215                 220
Ser Gln Ala Glu Thr Gly Glu Ile Lys Gly His Tyr Leu Asn Val Thr
225                 230                 235                 240
Ala Pro Thr Cys Glu Glu Met Met Lys Arg Ala Glu Phe Ala Lys Glu
                245                 250                 255
Leu Gly Met Pro Ile Ile Met His Asp Phe Leu Thr Ala Gly Phe Thr
            260                 265                 270
Ala Asn Thr Thr Leu Ala Lys Trp Cys Arg Asp Asn Gly Val Leu Leu
        275                 280                 285
His Ile His Arg Ala Gly His Ala Thr Ile Asp Arg Gln Arg Asn His
    290                 295                 300
Gly Ile His Phe Arg Val Leu Ala Lys Cys Leu Arg Leu Ser Gly Gly
305                 310                 315                 320
Asp His Leu His Ser Gly Thr Val Val Gly Lys Leu Glu Gly Asp Lys
                325                 330                 335
Ala Ser Thr Leu Gly Phe Val Asp Leu Met Arg Glu Asp His Ile Glu
            340                 345                 350
Ala Asp Arg Ser Arg Gly Val Phe Phe Thr Gln Asp Trp Ala Ser Met
        355                 360                 365
Pro Gly Val Leu Pro Val Ala Ser Gly Gly Ile His Val Trp His Met
    370                 375                 380
Pro Ala Leu Val Glu Ile Phe Gly Asp Asp Ser Val Leu Gln Phe Gly
385                 390                 395                 400
Gly Gly Thr Leu Gly His Pro Trp Gly Asn Ala Pro Gly Ala Thr Ala
                405                 410                 415
Asn Arg Val Ala Leu Glu Ala Cys Val Gln Ala Arg Asn Glu Gly Arg
            420                 425                 430
Asp Leu Tyr Arg Glu Gly Gly Asp Ile Leu Arg Glu Ala Gly Lys Trp
        435                 440                 445
Ser Pro Glu Leu Ala Ala Ala Leu Asp Leu Trp Lys Glu Ile Lys Phe
    450                 455                 460
Glu Phe Glu Thr Met Asp Lys Leu
465                 470
```

<210> 45
<211> 1845
<212> DNA
<213> Artificial

<220>
<223> Com base no nucleotídeo PCC6301 rbcLS de Synechococcus Sp. do tipo selvagem

<400> 45

```
atggccaaga cgcaatctgc cgcaggctat aaggccgggg tgaaggacta caaactcacc     60
tattacaccc ccgattacac ccccaaagac actgacatac tggcggcttt ccgcttcagc    120
cctcagccgg gtgtccctgc tgacgaagct ggtgcggcga tcgcggctga atcttcgacc    180
ggtacctgga ccaccgtgtg gaccgacttg ctgaccgaca tggatcggta caaaggcaag    240
tgctaccaca tcgagccggt gcaaggcgaa gagaactcct actttgcgtt catcgcttac    300
ccgctcgacc tgtttgaaga agggtcggtc accaacctcc tgacctcgat cgtcggtaac    360
```

```
gtgtttggct tcaaagctat ccgttcgctg cgtctggaag acatccgctt gcccgtcgcc    420
ttggtcaaaa ccttccaagg tcctccccac ggtatccaag tcgagcgcga cctgctgaac    480
aagtacggcc gtccgatgct gggttgcacg atcaaaccaa aactcggtct gtcggcgaaa    540
aactacggtc gtgccgtcta cgaatgtctg cgcggcggtc tggacttcac caaagacgac    600
gaaaacatca actcgcagcc gttccaacgc tggcgcgatc gcttcctgtt tgtggctgat    660
gcaatccaca aatcgcaagc agaaaccggt gaaatcaaag gtcactacct gaacgtgacc    720
gcgccgacct gcgaagaaat gatgaaacgg gctgagttcg ctaaagaact cggcatgccg    780
atcatcatgc atgacttctt gacggctggt ttcaccgcca acaccacctt ggcaaaatgg    840
tgccgcgaca acggcgtcct gctgcacatc caccgtgcaa tgcacgcggt gatcgaccgt    900
cagcgtaacc acgggattca cttccgtgtc ttggccaagt gtttgcgtct gtccggtggt    960
gaccacctcc actccggcac cgtcgtcggc aaactggaag gcgacaaagc ttcgaccttg   1020
ggctttgttg acttgatgcg cgaagaccac atcgaagctg accgcagccg tggggtcttc   1080
ttcacccaag attgggcgtc gatgccgggc gtgctgccgg ttgcttccgg tggtatccac   1140
gtgtggcaca tgcccgcact ggtggaaatc ttcggtgatg actccgttct ccagttcggt   1200
ggcggcacct tgggtcaccc ctggggtaat gctcctggtg caaccgcgaa ccgtgttgcc   1260
ttggaagctt gcgtccaagc tcggaacgaa ggtcgcgacc tctaccgtga aggcggcgac   1320
atccttcgtg aagctggcaa gtggtcgcct gaactggctg ctgccctcga cctctggaaa   1380
gagatcaagt tcgaattcga aacgatggac aagctctaag gagcctctga ctatcgctgg   1440
gggagtgagc gttgctgcgt aaagctttct ccccagcctt tcgacttaac ctttcaggat   1500
ttctgaatca tgagcatgaa aactctgccc aaagagcgtc gtttcgagac tttctcgtac   1560
ctgcctcccc tcagcgatcg ccaaatcgct gcacaaatcg agtacatgat cgagcaaggc   1620
ttccaccct tgatcgagtt caacgagcac tcgaatccgg aagagttcta ctggacgatg   1680
tggaagctcc ccctgtttga ctgcaagagc cctcagcaag tcctcgatga agtgcgtgag   1740
tgccgcagcg aatacggtga ttgctacatc cgtgtcgctg gcttcgacaa catcaagcag   1800
tgccaaaccg tgagcttcat cgttcatcgt cccggccgat actaa                   1845
```

<210> 46
<211> 472
<212> PRT
<213> Artificial

<220>
<223> Com base na sequência traduzida de PCC6301 rbcLS de Synechococcus Sp. do tipo selvagem

<400> 46

```
Met Ala Lys Thr Gln Ser Ala Ala Gly Tyr Lys Ala Gly Val Lys Asp
1               5                  10                  15
Tyr Lys Leu Thr Tyr Tyr Thr Pro Asp Tyr Thr Pro Lys Asp Thr Asp
            20                  25                  30
Ile Leu Ala Ala Phe Arg Phe Ser Pro Gln Pro Gly Val Pro Ala Asp
        35                  40                  45
Glu Ala Gly Ala Ala Ile Ala Ala Glu Ser Ser Thr Gly Thr Trp Thr
    50                  55                  60
Thr Val Trp Thr Asp Leu Leu Thr Asp Met Asp Arg Tyr Lys Gly Lys
65                  70                  75                  80
Cys Tyr His Ile Glu Pro Val Gln Gly Glu Glu Asn Ser Tyr Phe Ala
                85                  90                  95
Phe Ile Ala Tyr Pro Leu Asp Leu Phe Glu Glu Gly Ser Val Thr Asn
            100                 105                 110
Leu Leu Thr Ser Ile Val Gly Asn Val Phe Gly Phe Lys Ala Ile Arg
        115                 120                 125
Ser Leu Arg Leu Glu Asp Ile Arg Leu Pro Val Ala Leu Val Lys Thr
    130                 135                 140
Phe Gln Gly Pro Pro His Gly Ile Gln Val Glu Arg Asp Leu Leu Asn
145                 150                 155                 160
Lys Tyr Gly Arg Pro Met Leu Gly Cys Thr Ile Lys Pro Lys Leu Gly
                165                 170                 175
Leu Ser Ala Lys Asn Tyr Gly Arg Ala Val Tyr Glu Cys Leu Arg Gly
            180                 185                 190
```

```
Gly Leu Asp Phe Thr Lys Asp Asp Glu Asn Ile Asn Ser Gln Pro Phe
        195                 200                 205
Gln Arg Trp Arg Asp Arg Phe Leu Phe Val Ala Asp Ala Ile His Lys
    210                 215                 220
Ser Gln Ala Glu Thr Gly Glu Ile Lys Gly His Tyr Leu Asn Val Thr
225                 230                 235                 240
Ala Pro Thr Cys Glu Glu Met Met Lys Arg Ala Glu Phe Ala Lys Glu
                245                 250                 255
Leu Gly Met Pro Ile Ile Met His Asp Phe Leu Thr Ala Gly Phe Thr
            260                 265                 270
Ala Asn Thr Thr Leu Ala Lys Trp Cys Arg Asp Asn Gly Val Leu Leu
        275                 280                 285
His Ile His Arg Ala Met His Ala Val Ile Asp Arg Gln Arg Asn His
    290                 295                 300
Gly Ile His Phe Arg Val Leu Ala Lys Cys Leu Arg Leu Ser Gly Gly
305                 310                 315                 320
Asp His Leu His Ser Gly Thr Val Val Gly Lys Leu Glu Gly Asp Lys
                325                 330                 335
Ala Ser Thr Leu Gly Phe Val Asp Leu Met Arg Glu Asp His Ile Glu
            340                 345                 350
Ala Asp Arg Ser Arg Gly Val Phe Phe Thr Gln Asp Trp Ala Ser Met
        355                 360                 365
Pro Gly Val Leu Pro Val Ala Ser Gly Gly Ile His Val Trp His Met
    370                 375                 380
Pro Ala Leu Val Glu Ile Phe Gly Asp Asp Ser Val Leu Gln Phe Gly
385                 390                 395                 400
Gly Gly Thr Leu Gly His Pro Trp Gly Asn Ala Pro Gly Ala Thr Ala
                405                 410                 415
Asn Arg Val Ala Leu Glu Ala Cys Val Gln Ala Arg Asn Glu Gly Arg
            420                 425                 430
Asp Leu Tyr Arg Glu Gly Gly Asp Ile Leu Arg Glu Ala Gly Lys Trp
        435                 440                 445
Ser Pro Glu Leu Ala Ala Ala Leu Asp Leu Trp Lys Glu Ile Lys Phe
    450                 455                 460
Glu Phe Glu Thr Met Asp Lys Leu
465                 470
```

<210> 47
<211> 1845
<212> DNA
<213> Artificial

<220>
<223> Com base no nucleotídeo PCC6301 rbcLS de Synechococcus Sp. do tipo selvagem

<400> 47

```
atggccaaga cgcaatctgc cgcaggctat aaggccgggg tgaaggacta caaactcacc     60
tattacaccc ccgattacac ccccaaagac actgacatac tggcggcttt ccgcttcagc    120
cctcagccgg gtgtccctgc tgacgaagct ggtgcggcga tcgcggctga atcttcgacc    180
ggtacctgga ccaccgtgtg gaccgacttg ctgaccgaca tggatcggta caaaggcaag    240
tgctaccaca tcgagccggt gcaaggcgaa gagaactcct actttgcgtt catcgcttac    300
ccgctcgacc tgtttgaaga agggtcggtc accaacctcc tgacctcgat catcggtaac    360
gtgtttggct tcaaagctat ccgttcgctg cgtctggaag acatccgctt gcccgtcgcc    420
ttggtcaaaa ccttccaagg tcctccccac ggtatccaag tcgagcgcga cctgctgaac    480
aagtacggcc gtccgatgct gggttgcacg atcaaaccaa aactcggtct gtcggcgaaa    540
aactacggtc gtgccgtcta cgaatgtctg cgcggcggtc tggacttcac caaagacgac    600
gaaaacatca actcgcagcc gttccaacgc tggcgcgatc gcttcctgtt tgtggctgat    660
gcaatccaca aatcgcaagc agaaaccggt gaaatcaaag gtcactacct gaacgtgacc    720
gcgccgacct gcgaagaaat gatgaaacgg gctgagttcg ctaaagaact cggcatgccg    780
atcatcatgc atgacttctt gacggctggt ttcaccgcca acaccacctt ggcaaaatgg    840
```

```
tgccgcgaca acggcgtcct gctgcacatc caccgtgcag ggcacgcgac gatcgaccgt    900
cagcgtaacc acgggattca cttccgtgtc ttggccaagt gtttgcgtct gtccggtggt    960
gaccacctcc actccggcac cgtcgtcggc aaactggaag gcgacaaagc ttcgaccttg   1020
ggctttgttg acttgatgcg cgaagaccac atcgaagctg accgcagccg tggggtcttc   1080
ttcacccaag attgggcgtc gatgccgggc gtgctgccgg ttgcttccgg tggtatccac   1140
gtgtggcaca tgcccgcact ggtggaaatc ttcggtgatg actccgttct ccagttcggt   1200
ggcggcacct tgggtcaccc ctggggtaat gctcctggtg caaccgcgaa ccgtgttgcc   1260
ttggaagctt gcgtccaagc tcggaacgaa ggtcgcgacc tctaccgtga aggcggcgac   1320
atccttcgtg aagctggcaa gtggtcgcct gaactggctg ctgccctcga cctctggaaa   1380
gagatcaagt tcgaattcga aacgatggac aagctctaag gagcctctga ctatcgctgg   1440
gggagtgagc gttgctgcgt aaagctttct ccccagcctt tcgacttaac ctttcaggat   1500
ttctgaatca tgagcatgaa aactctgccc aaagagcgtc gtttcgagac tttctcgtac   1560
ctgcctcccc tcagcgatcg ccaaatcgct gcacaaatcg agtacatgat cgagcaaggc   1620
ttccaccccc tgatcgagtt caacgagcac tcgaatccgg aagagttcta ctggacgatg   1680
tggaagctcc ccctgtttga ctgcaagagc cctcagcaag tcctcgatga agtgcgtgag   1740
tgccgcagcg aatacggtga ttgctacatc cgtgtcgctg gcttcgacaa catcaagcag   1800
tgccaaaccg tgagcttcat cgttcatcgt cccggccgat actaa                   1845
```

<210> 48
<211> 472
<212> PRT
<213> Artificial

<220>
<223> Com base na sequência traduzida de PCC6301 rbcLS de Synechococcus Sp. do tipo selvagem

<400> 48

```
Met Ala Lys Thr Gln Ser Ala Ala Gly Tyr Lys Ala Gly Val Lys Asp
1               5                  10                  15
Tyr Lys Leu Thr Tyr Tyr Thr Pro Asp Tyr Thr Pro Lys Asp Thr Asp
            20                  25                  30
Ile Leu Ala Ala Phe Arg Phe Ser Pro Gln Pro Gly Val Pro Ala Asp
        35                  40                  45
Glu Ala Gly Ala Ala Ile Ala Ala Glu Ser Ser Thr Gly Thr Trp Thr
    50                  55                  60
Thr Val Trp Thr Asp Leu Leu Thr Asp Met Asp Arg Tyr Lys Gly Lys
65                  70                  75                  80
Cys Tyr His Ile Glu Pro Val Gln Gly Glu Glu Asn Ser Tyr Phe Ala
                85                  90                  95
Phe Ile Ala Tyr Pro Leu Asp Leu Phe Glu Glu Gly Ser Val Thr Asn
            100                 105                 110
Leu Leu Thr Ser Ile Ile Gly Asn Val Phe Gly Phe Lys Ala Ile Arg
        115                 120                 125
Ser Leu Arg Leu Glu Asp Ile Arg Leu Pro Val Ala Leu Val Lys Thr
    130                 135                 140
Phe Gln Gly Pro Pro His Gly Ile Gln Val Glu Arg Asp Leu Leu Asn
145                 150                 155                 160
Lys Tyr Gly Arg Pro Met Leu Gly Cys Thr Ile Lys Pro Lys Leu Gly
                165                 170                 175
Leu Ser Ala Lys Asn Tyr Gly Arg Ala Val Tyr Glu Cys Leu Arg Gly
            180                 185                 190
Gly Leu Asp Phe Thr Lys Asp Asp Glu Asn Ile Asn Ser Gln Pro Phe
        195                 200                 205
Gln Arg Trp Arg Asp Arg Phe Leu Phe Val Ala Asp Ala Ile His Lys
    210                 215                 220
Ser Gln Ala Glu Thr Gly Glu Ile Lys Gly His Tyr Leu Asn Val Thr
225                 230                 235                 240
Ala Pro Thr Cys Glu Glu Met Met Lys Arg Ala Glu Phe Ala Lys Glu
                245                 250                 255
```

```
Leu Gly Met Pro Ile Ile Met His Asp Phe Leu Thr Ala Gly Phe Thr
            260                 265                 270
Ala Asn Thr Thr Leu Ala Lys Trp Cys Arg Asp Asn Gly Val Leu Leu
        275                 280                 285
His Ile His Arg Ala Gly His Ala Thr Ile Asp Arg Gln Arg Asn His
    290                 295                 300
Gly Ile His Phe Arg Val Leu Ala Lys Cys Leu Arg Leu Ser Gly Gly
305                 310                 315                 320
Asp His Leu His Ser Gly Thr Val Val Gly Lys Leu Glu Gly Asp Lys
                325                 330                 335
Ala Ser Thr Leu Gly Phe Val Asp Leu Met Arg Glu Asp His Ile Glu
            340                 345                 350
Ala Asp Arg Ser Arg Gly Val Phe Phe Thr Gln Asp Trp Ala Ser Met
        355                 360                 365
Pro Gly Val Leu Pro Val Ala Ser Gly Gly Ile His Val Trp His Met
    370                 375                 380
Pro Ala Leu Val Glu Ile Phe Gly Asp Asp Ser Val Leu Gln Phe Gly
385                 390                 395                 400
Gly Gly Thr Leu Gly His Pro Trp Gly Asn Ala Pro Gly Ala Thr Ala
                405                 410                 415
Asn Arg Val Ala Leu Glu Ala Cys Val Gln Ala Arg Asn Glu Gly Arg
            420                 425                 430
Asp Leu Tyr Arg Glu Gly Gly Asp Ile Leu Arg Glu Ala Gly Lys Trp
        435                 440                 445
Ser Pro Glu Leu Ala Ala Ala Leu Asp Leu Trp Lys Glu Ile Lys Phe
    450                 455                 460
Glu Phe Glu Thr Met Asp Lys Leu
465                 470
```

<210> 49
<211> 1845
<212> DNA
<213> Artificial

<220>
<223> Com base no nucleotídeo PCC6301 rbcLS de Synechococcus Sp. do tipo selvagem

<400> 49

```
atggccaaga cgcaatctgc cgcaggctat aaggccgggg tgatagacta caaactcggc     60
tattacaccc ccgattacac ccccaaagac actgacctgc tggcggcttt ccgcttcagc    120
cctcagccgg gtgtccctgc tgacgaagct ggtgcggcga tcgcggctga atcttcgacc    180
ggtacctgga ccaccgtgtg gaccgacttg ctgaccgaca tggatcggta caaaggcaag    240
tgctaccaca tcgagccggt gcaaggcgaa gagaactcct actttgcgtt catcgcttac    300
ccgctcgacc tgtttgaaga agggtcggtc accaacatcc tgacctcgat cgtcggtaac    360
gtgtttggct tcaaagctat ccgttcgctg cgtctggaag acatccgctt ccccgtcgcc    420
ttggtcaaaa ccttccaagg tcctccccac ggtatccaag tcgagcgcga cctgctgaac    480
aagtacggcc gtccgatgct gggttgcacg atcaaaccaa aactcggtct gtcggcgaaa    540
aactacggtc gtgccgtcta cgaatgtctg cgcggcggtc tggacttcac caaagacgac    600
gaaaacatca actcgcagcc gttccaacgc tggcgcgatc gcttcctgtt tgtggctgat    660
gcaatccaca aatcgcaagc agaaaccggt gaaatcaaag gtcactacct gaacgtgacc    720
gcgccgacct gcgaagaaat gatgaaacgg gctgagttcg ctaaagaact cggcatgccg    780
atcatcatgc atgacttctt gacggctggt ttcaccgcca acaccacctt ggcaaaatgg    840
tgccgcgaca acggcgtcct gctgcacatc caccgtgcaa tgcacgcggt gatcgaccgt    900
cagcgtaacc acgggattca cttccgtgtc ttggccaagt gtttgcgtct gtccggtggt    960
gaccacctcc actccggcac cgtcgtcggc aaactggaag gcgacaaagc ttcgaccttg   1020
ggctttgttg acttgatgcg cgaagaccac atcgaagctg accgcagccg tggggtcttc   1080
ttcacccaag attgggcgtc gatgccgggc gtgctgccgg ttgcttccgg tggtatccac   1140
gtgtggcaca tgcccgcact ggtggaaatc ttcggtgatg actccgttct ccagttcggt   1200
ggcggcacct tgggtcaccc ctggggtaat gctcctggtg caaccgcgaa ccgtgttgcc   1260
ttggaagctt gcgtccaagc tcggaacgaa ggtcgcgacc tctaccgtga aggcggcgac   1320
```

```
atccttcgtg aagctggcaa gtggtcgcct gaactggctg ctgccctcga cctctggaaa   1380
gagatcaagt tcgaattcga aacgatggac aagctctaag gagcctctga ctatcgctgg   1440
gggagtgagc gttgctgcgt aaagctttct ccccagcctt tcgacttaac ctttcaggat   1500
ttctgaatca tgagcatgaa aactctgccc aaagagcgtc gtttcgagac tttctcgtac   1560
ctgcctcccc tcagcgatcg ccaaatcgct gcacaaatcg agtacatgat cgagcaaggc   1620
ttccaccccc tgatcgagtt caacgagcac tcgaatccgg aagagttcta ctggacgatg   1680
tggaagctcc ccctgtttga ctgcaagagc cctcagcaag tcctcgatga agtgcgtgag   1740
tgccgcagcg aatacggtga ttgctacatc cgtgtcgctg gcttcgacaa catcaagcag   1800
tgccaaaccg tgagcttcat cgttcatcgt cccggccgat actaa                   1845
```

<210> 50
<211> 472
<212> PRT
<213> Artificial

<220>
<223> Com base na sequência traduzida de PCC6301 rbcLS de Synechococcus Sp. do tipo selvagem

<400> 50

```
Met Ala Lys Thr Gln Ser Ala Ala Gly Tyr Lys Ala Gly Val Ile Asp
1               5                  10                  15
Tyr Lys Leu Gly Tyr Tyr Thr Pro Asp Tyr Thr Pro Lys Asp Thr Asp
            20                  25                  30
Leu Leu Ala Ala Phe Arg Phe Ser Pro Gln Pro Gly Val Pro Ala Asp
        35                  40                  45
Glu Ala Gly Ala Ala Ile Ala Ala Glu Ser Ser Thr Gly Thr Trp Thr
    50                  55                  60
Thr Val Trp Thr Asp Leu Leu Thr Asp Met Asp Arg Tyr Lys Gly Lys
65                  70                  75                  80
Cys Tyr His Ile Glu Pro Val Gln Gly Glu Glu Asn Ser Tyr Phe Ala
                85                  90                  95
Phe Ile Ala Tyr Pro Leu Asp Leu Phe Glu Glu Gly Ser Val Thr Asn
            100                 105                 110
Ile Leu Thr Ser Ile Val Gly Asn Val Phe Gly Phe Lys Ala Ile Arg
        115                 120                 125
Ser Leu Arg Leu Glu Asp Ile Arg Phe Pro Val Ala Leu Val Lys Thr
    130                 135                 140
Phe Gln Gly Pro Pro His Gly Ile Gln Val Glu Arg Asp Leu Leu Asn
145                 150                 155                 160
Lys Tyr Gly Arg Pro Met Leu Gly Cys Thr Ile Lys Pro Lys Leu Gly
                165                 170                 175
Leu Ser Ala Lys Asn Tyr Gly Arg Ala Val Tyr Glu Cys Leu Arg Gly
            180                 185                 190
Gly Leu Asp Phe Thr Lys Asp Asp Glu Asn Ile Asn Ser Gln Pro Phe
        195                 200                 205
Gln Arg Trp Arg Asp Arg Phe Leu Phe Val Ala Asp Ala Ile His Lys
    210                 215                 220
Ser Gln Ala Glu Thr Gly Glu Ile Lys Gly His Tyr Leu Asn Val Thr
225                 230                 235                 240
Ala Pro Thr Cys Glu Glu Met Met Lys Arg Ala Glu Phe Ala Lys Glu
                245                 250                 255
Leu Gly Met Pro Ile Ile Met His Asp Phe Leu Thr Ala Gly Phe Thr
            260                 265                 270
Ala Asn Thr Thr Leu Ala Lys Trp Cys Arg Asp Asn Gly Val Leu Leu
        275                 280                 285
His Ile His Arg Ala Met His Ala Val Ile Asp Arg Gln Arg Asn His
    290                 295                 300
Gly Ile His Phe Arg Val Leu Ala Lys Cys Leu Arg Leu Ser Gly Gly
305                 310                 315                 320
```

```
Asp His Leu His Ser Gly Thr Val Val Gly Lys Leu Glu Gly Asp Lys
                325                 330                 335
Ala Ser Thr Leu Gly Phe Val Asp Leu Met Arg Glu Asp His Ile Glu
            340                 345                 350
Ala Asp Arg Ser Arg Gly Val Phe Phe Thr Gln Asp Trp Ala Ser Met
        355                 360                 365
Pro Gly Val Leu Pro Val Ala Ser Gly Gly Ile His Val Trp His Met
    370                 375                 380
Pro Ala Leu Val Glu Ile Phe Gly Asp Asp Ser Val Leu Gln Phe Gly
385                 390                 395                 400
Gly Gly Thr Leu Gly His Pro Trp Gly Asn Ala Pro Gly Ala Thr Ala
                405                 410                 415
Asn Arg Val Ala Leu Glu Ala Cys Val Gln Ala Arg Asn Glu Gly Arg
            420                 425                 430
Asp Leu Tyr Arg Glu Gly Gly Asp Ile Leu Arg Glu Ala Gly Lys Trp
        435                 440                 445
Ser Pro Glu Leu Ala Ala Ala Leu Asp Leu Trp Lys Glu Ile Lys Phe
    450                 455                 460
Glu Phe Glu Thr Met Asp Lys Leu
465                 470
```

<210> 51
<211> 1848
<212> DNA
<213> Artificial

<220>
<223> Com base no nucleotídeo PCC6301 rbcLS de Synechococcus Sp. do tipo selvagem

<400> 51

```
atggccaaga cgcaatctgc cgcaggctat aaggccgggg tgaaggacta cgctaagatg     60
ggatactgga cccccgatta caccccaaa gacactgacc tgctggcggc tttccgcttc    120
agccctcagc cgggtgtccc tgctgacgaa gctggtgcgg cgatcgcggc tgaatcttcg    180
accggtacct ggaccaccgt gtggaccgac ttgctgaccg acatggatcg gtacaaaggc    240
aagtgctacc acatcgagcc ggtgcaaggc gaagagaact cctactttgc gttcatcgct    300
tacccgctcg acctgtttga agaagggtcg gtcaccaaca tcctgacctc gatcgtcggt    360
aacgtgtttg gcttcaaagc tatccgttcg ctgcgtctgg aagacatccg cttccccgtc    420
gccttggtca aaaccttcca aggtcctccc cacggtatcc aagtcgagcg cgacctgctg    480
aacaagtacg gccgtccgat gctgggttgc acgatcaaac caaaactcgg tctgtcggcg    540
aaaaactacg gtcgtgccgt ctacgaatgt ctgcgcggcg gtctggactt caccaaagac    600
gacgaaaaca tcaactcgca gccgttccaa cgctggcgcg atcgcttcct gtttgtggct    660
gatgcaatcc acaaatcgca agcagaaacc ggtgaaatca aaggtcacta cctgaacgtg    720
accgcgccga cctgcgaaga aatgatgaaa cgggctgagt tcgctaaaga actcggcatg    780
ccgatcatca tgcatgactt cttgacggct ggtttcaccg ccaacaccac cttggcaaaa    840
tggtgccgcg acaacggcgt cctgctgcac atccaccgtg caatgcacgc ggtgatcgac    900
cgtcagcgta accacgggat tcacttccgt gtcttggcca agtgtttgcg tctgtccggt    960
ggtgaccacc tccactccgg caccgtcgtc ggcaaactgg aaggcgacaa agcttcgacc   1020
ttgggctttg ttgacttgat gcgcgaagac cacatcgaag ctgaccgcag ccgtggggtc   1080
ttcttcaccc aagattgggc gtcgatgccg ggcgtgctgc cggttgcttc cggtggtatc   1140
cacgtgtggc acatgcccgc actggtggaa atcttcggtg atgactccgt tctccagttc   1200
ggtggcggca ccttgggtca cccctggggt aatgctcctg gtgcaaccgc gaaccgtgtt   1260
gccttggaag cttgcgtcca agctcggaac gaaggtcgcg acctctaccg tgaaggcggc   1320
gacatccttc gtgaagctgg caagtggtcg cctgaactgg ctgctgccct cgacctctgg   1380
aaagagatca agttcgaatt cgaaacgatg gacaagctct aaggagcctc tgactatcgc   1440
tgggggagtg agcgttgctg cgtaaagctt tctccccagc ctttcgactt aacctttcag   1500
gatttctgaa tcatgagcat gaaaactctg cccaaagagc gtcgtttcga gactttctcg   1560
tacctgcctc ccctcagcga tcgccaaatc gctgcacaaa tcgagtacat gatcgagcaa   1620
ggcttccacc ccttgatcga gttcaacgag cactcgaatc cggaagagtt ctactggacg   1680
atgtggaagc tccccctgtt tgactgcaag agccctcagc aagtcctcga tgaagtgcgt   1740
gagtgccgca gcgaatacgg tgattgctac atccgtgtcg ctggcttcga caacatcaag   1800
```

cagtgccaaa ccgtgagctt catcgttcat cgtcccggcc gatactaa 1848

<210> 52
<211> 473
<212> PRT
<213> Artificial

<220>
<223> Com base na sequência traduzida de PCC6301 rbcLS de Synechococcus Sp. do tipo selvagem

<400> 52

```
Met Ala Lys Thr Gln Ser Ala Ala Gly Tyr Lys Ala Gly Val Lys Asp
1               5                  10                  15
Tyr Ala Lys Met Gly Tyr Trp Thr Pro Asp Tyr Thr Pro Lys Asp Thr
            20                  25                  30
Asp Leu Leu Ala Ala Phe Arg Phe Ser Pro Gln Pro Gly Val Pro Ala
        35                  40                  45
Asp Glu Ala Gly Ala Ala Ile Ala Ala Glu Ser Ser Thr Gly Thr Trp
    50                  55                  60
Thr Thr Val Trp Thr Asp Leu Leu Thr Asp Met Asp Arg Tyr Lys Gly
65                  70                  75                  80
Lys Cys Tyr His Ile Glu Pro Val Gln Gly Glu Glu Asn Ser Tyr Phe
                85                  90                  95
Ala Phe Ile Ala Tyr Pro Leu Asp Leu Phe Glu Glu Gly Ser Val Thr
            100                 105                 110
Asn Ile Leu Thr Ser Ile Val Gly Asn Val Phe Gly Phe Lys Ala Ile
        115                 120                 125
Arg Ser Leu Arg Leu Glu Asp Ile Arg Phe Pro Val Ala Leu Val Lys
    130                 135                 140
Thr Phe Gln Gly Pro Pro His Gly Ile Gln Val Glu Arg Asp Leu Leu
145                 150                 155                 160
Asn Lys Tyr Gly Arg Pro Met Leu Gly Cys Thr Ile Lys Pro Lys Leu
                165                 170                 175
Gly Leu Ser Ala Lys Asn Tyr Gly Arg Ala Val Tyr Glu Cys Leu Arg
            180                 185                 190
Gly Gly Leu Asp Phe Thr Lys Asp Asp Glu Asn Ile Asn Ser Gln Pro
        195                 200                 205
Phe Gln Arg Trp Arg Asp Arg Phe Leu Phe Val Ala Asp Ala Ile His
    210                 215                 220
Lys Ser Gln Ala Glu Thr Gly Glu Ile Lys Gly His Tyr Leu Asn Val
225                 230                 235                 240
Thr Ala Pro Thr Cys Glu Glu Met Met Lys Arg Ala Glu Phe Ala Lys
                245                 250                 255
Glu Leu Gly Met Pro Ile Ile Met His Asp Phe Leu Thr Ala Gly Phe
            260                 265                 270
Thr Ala Asn Thr Thr Leu Ala Lys Trp Cys Arg Asp Asn Gly Val Leu
        275                 280                 285
Leu His Ile His Arg Ala Met His Ala Val Ile Asp Arg Gln Arg Asn
    290                 295                 300
His Gly Ile His Phe Arg Val Leu Ala Lys Cys Leu Arg Leu Ser Gly
305                 310                 315                 320
Gly Asp His Leu His Ser Gly Thr Val Val Gly Lys Leu Glu Gly Asp
                325                 330                 335
Lys Ala Ser Thr Leu Gly Phe Val Asp Leu Met Arg Glu Asp His Ile
            340                 345                 350
Glu Ala Asp Arg Ser Arg Gly Val Phe Phe Thr Gln Asp Trp Ala Ser
        355                 360                 365
Met Pro Gly Val Leu Pro Val Ala Ser Gly Gly Ile His Val Trp His
    370                 375                 380
```

```
Met Pro Ala Leu Val Glu Ile Phe Gly Asp Asp Ser Val Leu Gln Phe
385                 390                 395                 400
Gly Gly Gly Thr Leu Gly His Pro Trp Gly Asn Ala Pro Gly Ala Thr
                405                 410                 415
Ala Asn Arg Val Ala Leu Glu Ala Cys Val Gln Ala Arg Asn Glu Gly
            420                 425                 430
Arg Asp Leu Tyr Arg Glu Gly Gly Asp Ile Leu Arg Glu Ala Gly Lys
        435                 440                 445
Trp Ser Pro Glu Leu Ala Ala Ala Leu Asp Leu Trp Lys Glu Ile Lys
    450                 455                 460
Phe Glu Phe Glu Thr Met Asp Lys Leu
465                 470
```

<210> 53
<211> 1845
<212> DNA
<213> Artificial

<220>
<223> Com base no nucleotídeo PCC6301 rbcLS de Synechococcus Sp. do tipo selvagem

<400> 53

```
atggccaaga cgcaatctgc cgcaggctat aaggccgggg tgatagacta caaactcggc     60
tattacaccc ccgattacac ccccaaagac actgacctgc tggcggcttt ccgcttcagc    120
cctcagccgg gtgtccctgc tgacgaagct ggtgcggcga tcgcggctga atcttcgacc    180
ggtacctgga ccgtcgtgtg gaccgacttg ctgaccgaca tggatcggta ccgcggcaag    240
tgctaccaca tcgagccggt gcaaggcgaa gagaactcct actttgcgtt catcgcttac    300
ccgctcgacc tgtttgaaga agggtcggtc accaacatcc tgacctcgat cgtcggtaac    360
gtgtttggct tcaaagctat ccgttcgctg cgtctggaag acatccgctt ccccgtcgcc    420
ttggtcaaaa ccttccaagg tcctccccac ggtatccaag tcgagcgcga cctgctgaac    480
aagtacggcc gtccgatgct gggttgcacg atcaaaccaa aactcggtct gtcggcgaaa    540
aactacggtc gtgccgtcta cgaatgtctg cgcggcggtc tggacttcac caaagacgac    600
gaaaacatca actcgcagcc gttccaacgc tggcgcgatc gcttcctgtt tgtggctgat    660
gcaatccaca aatcgcaagc agaaaccggt gaaatcaaag gtcactacct gaacgtgacc    720
gcgccgacct gcgaagaaat gatgaaacgg gctgagttcg ctaaagaact cggcatgccg    780
atcatcatgc atgacttctt gacggctggt ttcaccgcca acaccacctt ggcaaaatgg    840
tgccgcgaca acggcgtcct gctgcacatc caccgtgcaa tgcacgcggt gatcgaccgt    900
cagcgtaacc acgggattca cttccgtgtc ttggccaagt gtttgcgtct gtccggtggt    960
gaccacctcc actccggcac cgtcgtcggc aaactggaag gcgacaaagc ttcgaccttg   1020
ggctttgttg acttgatgcg cgaagaccac atcgaagctg accgcagccg tggggtcttc   1080
ttcacccaag attgggcgtc gatgccgggc gtgctgccgg ttgcttccgg tggtatccac   1140
gtgtggcaca tgcccgcact ggtggaaatc ttcggtgatg actccgttct ccagttcggt   1200
ggcggcacct tgggtcaccc ctggggtaat gctcctggtg caaccgcgaa ccgtgttgcc   1260
ttggaagctt gcgtccaagc tcggaacgaa ggtcgcgacc tctaccgtga aggcggcgac   1320
atccttcgtg aagctggcaa gtggtcgcct gaactggctg ctgccctcga cctctggaaa   1380
gagatcaagt tcgaattcga aacgatggac aagctctaag gagcctctga ctatcgctgg   1440
gggagtgagc gttgctgcgt aaagctttct ccccagcctt tcgacttaac ctttcaggat   1500
ttctgaatca tgagcatgaa aactctgccc aaagagcgtc gtttcgagac tttctcgtac   1560
ctgcctcccc tcagcgatcg ccaaatcgct gcacaaatcg agtacatgat cgagcaaggc   1620
ttccaccccc tgatcgagtt caacgagcac tcgaatccgg aagagttcta ctggacgatg   1680
tggaagctcc ccctgtttga ctgcaagagc cctcagcaag tcctcgatga agtgcgtgag   1740
tgccgcagcg aatacggtga ttgctacatc cgtgtcgctg gcttcgacaa catcaagcag   1800
tgccaaaccg tgagcttcat cgttcatcgt cccggccgat actaa                   1845
```

<210> 54
<211> 472
<212> PRT
<213> Artificial

<220>

<223> Com base na sequência traduzida de PCC6301 rbcLS de Synechococcus Sp. do tipo selvagem

<400> 54

```
Met Ala Lys Thr Gln Ser Ala Ala Gly Tyr Lys Ala Gly Val Ile Asp
1               5                  10                  15
Tyr Lys Leu Gly Tyr Tyr Thr Pro Asp Tyr Thr Pro Lys Asp Thr Asp
            20                  25                  30
Leu Leu Ala Ala Phe Arg Phe Ser Pro Gln Pro Gly Val Pro Ala Asp
        35                  40                  45
Glu Ala Gly Ala Ala Ile Ala Ala Glu Ser Ser Thr Gly Thr Trp Thr
    50                  55                  60
Val Val Trp Thr Asp Leu Leu Thr Asp Met Asp Arg Tyr Arg Gly Lys
65                  70                  75                  80
Cys Tyr His Ile Glu Pro Val Gln Gly Glu Glu Asn Ser Tyr Phe Ala
                85                  90                  95
Phe Ile Ala Tyr Pro Leu Asp Leu Phe Glu Glu Gly Ser Val Thr Asn
            100                 105                 110
Ile Leu Thr Ser Ile Val Gly Asn Val Phe Gly Phe Lys Ala Ile Arg
        115                 120                 125
Ser Leu Arg Leu Glu Asp Ile Arg Phe Pro Val Ala Leu Val Lys Thr
    130                 135                 140
Phe Gln Gly Pro Pro His Gly Ile Gln Val Glu Arg Asp Leu Leu Asn
145                 150                 155                 160
Lys Tyr Gly Arg Pro Met Leu Gly Cys Thr Ile Lys Pro Lys Leu Gly
                165                 170                 175
Leu Ser Ala Lys Asn Tyr Gly Arg Ala Val Tyr Glu Cys Leu Arg Gly
            180                 185                 190
Gly Leu Asp Phe Thr Lys Asp Asp Glu Asn Ile Asn Ser Gln Pro Phe
        195                 200                 205
Gln Arg Trp Arg Asp Arg Phe Leu Phe Val Ala Asp Ala Ile His Lys
    210                 215                 220
Ser Gln Ala Glu Thr Gly Glu Ile Lys Gly His Tyr Leu Asn Val Thr
225                 230                 235                 240
Ala Pro Thr Cys Glu Glu Met Met Lys Arg Ala Glu Phe Ala Lys Glu
                245                 250                 255
Leu Gly Met Pro Ile Ile Met His Asp Phe Leu Thr Ala Gly Phe Thr
            260                 265                 270
Ala Asn Thr Thr Leu Ala Lys Trp Cys Arg Asp Asn Gly Val Leu Leu
        275                 280                 285
His Ile His Arg Ala Met His Ala Val Ile Asp Arg Gln Arg Asn His
    290                 295                 300
Gly Ile His Phe Arg Val Leu Ala Lys Cys Leu Arg Leu Ser Gly Gly
305                 310                 315                 320
Asp His Leu His Ser Gly Thr Val Val Gly Lys Leu Glu Gly Asp Lys
                325                 330                 335
Ala Ser Thr Leu Gly Phe Val Asp Leu Met Arg Glu Asp His Ile Glu
            340                 345                 350
Ala Asp Arg Ser Arg Gly Val Phe Phe Thr Gln Asp Trp Ala Ser Met
        355                 360                 365
Pro Gly Val Leu Pro Val Ala Ser Gly Gly Ile His Val Trp His Met
    370                 375                 380
Pro Ala Leu Val Glu Ile Phe Gly Asp Asp Ser Val Leu Gln Phe Gly
385                 390                 395                 400
Gly Gly Thr Leu Gly His Pro Trp Gly Asn Ala Pro Gly Ala Thr Ala
                405                 410                 415
Asn Arg Val Ala Leu Glu Ala Cys Val Gln Ala Arg Asn Glu Gly Arg
            420                 425                 430
Asp Leu Tyr Arg Glu Gly Gly Asp Ile Leu Arg Glu Ala Gly Lys Trp
        435                 440                 445
```

```
Ser Pro Glu Leu Ala Ala Ala Leu Asp Leu Trp Lys Glu Ile Lys Phe
    450                 455                 460
Glu Phe Glu Thr Met Asp Lys Leu
465                 470
```

<210> 55
<211> 1845
<212> DNA
<213> Artificial

<220>
<223> Com base no nucleotídeo PCC6301 rbcLS de Synechococcus Sp. do tipo selvagem

<400> 55

```
atggccaaga cgcaatctgc cgcaggctat aaggccgggg tgaaggacta caaactcggc     60
tattacaccc ccgattacac ccccaaagac actgacctgc tggcggcttt ccgcttcagc    120
cctcagccgg gtgtccctgc tgacgaagct ggtgcggcga tcgcggctga atcttcgacc    180
ggtacctgga ccaccgtgtg gaccgacttg ctgaccgaca tggatcggta ccgcggcaag    240
tgctaccaca tcgagccggt gcaaggcgaa gagaactcct actttgcgtt catcgcttac    300
gagctcgacc tgtttgaaga agggtcggtc accaacatcc tgacctcgat cgtcggtaac    360
gtgtttggct tcaaagctat ccgttcgctg cgtctggaag acatccgctt ccccgtcgcc    420
ttggtcaaaa ccttccaagg tcctccccac ggtatccaag tcgagcgcga cctgctgaac    480
aagtacggcc gtccgatgct gggttgcacg atcaaaccaa aactcggtct gtcggcgaaa    540
aactacggtc gtgccgtcta cgaatgtctg cgcggcggtc tggacttcac caaagacgac    600
gaaaacatca actcgcagcc gttccaacgc tggcgcgatc gcttcctgtt tgtggctgat    660
gcaatccaca aatcgcaagc agaaaccggt gaaatcaaag gtcactacct gaacgtgacc    720
gcgccgacct gcgaagaaat gatgaaacgg gctgagttcg ctaaagaact cggcatgccg    780
atcatcatgc atgacttctt gacggctggt ttcaccgcca acaccacctt ggcaaaatgg    840
tgccgcgaca acggcgtcct gctgcacatc caccgtgcaa tgcacgcggt gatcgaccgt    900
cagcgtaacc acgggattca cttccgtgtc ttggccaagt gtttgcgtct gtccggtggt    960
gaccacctcc actccggcac cgtcgtcggc aaactggaag gcgacaaagc ttcgaccttg   1020
ggctttgttg acttgatgcg cgaagaccac atcgaagctg accgcagccg tggggtcttc   1080
ttcacccaag attgggcgtc gatgccgggc gtgctgccgg ttgcttccgg tggtatccac   1140
gtgtggcaca tgcccgcact ggtggaaatc ttcggtgatg actccgttct ccagttcggt   1200
ggcggcacct tgggtcaccc ctggggtaat gctcctggtg caaccgcgaa ccgtgttgcc   1260
ttggaagctt gcgtccaagc tcggaacgaa ggtcgcgacc tctaccgtga aggcggcgac   1320
atccttcgtg aagctggcaa gtggtcgcct gaactggctg ctgccctcga cctctggaaa   1380
gagatcaagt tcgaattcga aacgatggac aagctctaag gagcctctga ctatcgctgg   1440
gggagtgagc gttgctgcgt aaagctttct ccccagcctt tcgacttaac ctttcaggat   1500
ttctgaatca tgagcatgaa aactctgccc aaagagcgtc gtttcgagac tttctcgtac   1560
ctgcctcccc tcagcgatcg ccaaatcgct gcacaaatcg agtacatgat cgagcaaggc   1620
ttccaccccc tgatcgagtt caacgagcac tcgaatccgg aagagttcta ctggacgatg   1680
tggaagctcc ccctgtttga ctgcaagagc cctcagcaag tcctcgatga agtgcgtgag   1740
tgccgcagcg aatacggtga ttgctacatc cgtgtcgctg gcttcgacaa catcaagcag   1800
tgccaaaccg tgagcttcat cgttcatcgt cccggccgat actaa                   1845
```

<210> 56
<211> 472
<212> PRT
<213> Artificial

<220>
<223> Com base na sequência traduzida de PCC6301 rbcLS de Synechococcus Sp. do tipo selvagem

<400> 56

```
Met Ala Lys Thr Gln Ser Ala Ala Gly Tyr Lys Ala Gly Val Lys Asp
1               5                  10                  15
Tyr Lys Leu Gly Tyr Tyr Thr Pro Asp Tyr Thr Pro Lys Asp Thr Asp
```

```
            20                  25                  30
Leu Leu Ala Ala Phe Arg Phe Ser Pro Gln Pro Gly Val Pro Ala Asp
        35                  40                  45
Glu Ala Gly Ala Ala Ile Ala Ala Glu Ser Ser Thr Gly Thr Trp Thr
    50                  55                  60
Thr Val Trp Thr Asp Leu Leu Thr Asp Met Asp Arg Tyr Arg Gly Lys
65                  70                  75                  80
Cys Tyr His Ile Glu Pro Val Gln Gly Glu Glu Asn Ser Tyr Phe Ala
                85                  90                  95
Phe Ile Ala Tyr Glu Leu Asp Leu Phe Glu Glu Gly Ser Val Thr Asn
            100                 105                 110
Ile Leu Thr Ser Ile Val Gly Asn Val Phe Gly Phe Lys Ala Ile Arg
        115                 120                 125
Ser Leu Arg Leu Glu Asp Ile Arg Phe Pro Val Ala Leu Val Lys Thr
    130                 135                 140
Phe Gln Gly Pro Pro His Gly Ile Gln Val Glu Arg Asp Leu Leu Asn
145                 150                 155                 160
Lys Tyr Gly Arg Pro Met Leu Gly Cys Thr Ile Lys Pro Lys Leu Gly
                165                 170                 175
Leu Ser Ala Lys Asn Tyr Gly Arg Ala Val Tyr Glu Cys Leu Arg Gly
            180                 185                 190
Gly Leu Asp Phe Thr Lys Asp Asp Glu Asn Ile Asn Ser Gln Pro Phe
        195                 200                 205
Gln Arg Trp Arg Asp Arg Phe Leu Phe Val Ala Asp Ala Ile His Lys
    210                 215                 220
Ser Gln Ala Glu Thr Gly Glu Ile Lys Gly His Tyr Leu Asn Val Thr
225                 230                 235                 240
Ala Pro Thr Cys Glu Glu Met Met Lys Arg Ala Glu Phe Ala Lys Glu
                245                 250                 255
Leu Gly Met Pro Ile Ile Met His Asp Phe Leu Thr Ala Gly Phe Thr
            260                 265                 270
Ala Asn Thr Thr Leu Ala Lys Trp Cys Arg Asp Asn Gly Val Leu Leu
        275                 280                 285
His Ile His Arg Ala Met His Ala Val Ile Asp Arg Gln Arg Asn His
    290                 295                 300
Gly Ile His Phe Arg Val Leu Ala Lys Cys Leu Arg Leu Ser Gly Gly
305                 310                 315                 320
Asp His Leu His Ser Gly Thr Val Val Gly Lys Leu Glu Gly Asp Lys
                325                 330                 335
Ala Ser Thr Leu Gly Phe Val Asp Leu Met Arg Glu Asp His Ile Glu
            340                 345                 350
Ala Asp Arg Ser Arg Gly Val Phe Phe Thr Gln Asp Trp Ala Ser Met
        355                 360                 365
Pro Gly Val Leu Pro Val Ala Ser Gly Gly Ile His Val Trp His Met
    370                 375                 380
Pro Ala Leu Val Glu Ile Phe Gly Asp Asp Ser Val Leu Gln Phe Gly
385                 390                 395                 400
Gly Gly Thr Leu Gly His Pro Trp Gly Asn Ala Pro Gly Ala Thr Ala
                405                 410                 415
Asn Arg Val Ala Leu Glu Ala Cys Val Gln Ala Arg Asn Glu Gly Arg
            420                 425                 430
Asp Leu Tyr Arg Glu Gly Gly Asp Ile Leu Arg Glu Ala Gly Lys Trp
        435                 440                 445
Ser Pro Glu Leu Ala Ala Ala Leu Asp Leu Trp Lys Glu Ile Lys Phe
    450                 455                 460
Glu Phe Glu Thr Met Asp Lys Leu
465                 470
```

<210> 57
<211> 1845
<212> DNA

<213> Artificial

<220>
<223> Com base no nucleotídeo PCC6301 rbcLS de Synechococcus Sp. do tipo selvagem

<400> 57
atggccaaga cgcaatctgc cgcaggctat aaggccgggg tgaagcccta caaactcggc 60
tattacaccc ccgattacac ccccaaagac actgacctgc tggcggcttt ccgcttcagc 120
cctcagccgg gtgtccctgc tgacgaagct ggtgcggcga tcgcggctga atcttcgacc 180
ggtacctgga ccaccgtgtg gaccgacttg ctgaccgaca tggatcggta ccgcggcaag 240
tgctaccaca tcgagccggt gcaaggcgaa gagaactcct actttgcgtt catcgcttac 300
ccgctcgacc tgtttgaaga agggtcggtc accaacatcc tgacctcgat cgtcggtaac 360
gtgtttggct tcaaagctat ccgttcgctg cgtctggaag acatccgctt ccccgtcgcc 420
ttggtcaaaa ccttccaagg tcctccccac ggtatccaag tcgagcgcga cctgctgaac 480
aagtacggcc gtccgatgct gggttgcacg atcaaaccaa aactcggtct gtcggcgaaa 540
aactacggtc gtgccgtcta cgaatgtctg cgcggcggtc tggacttcac caaagacgac 600
gaaaacatca actcgcagcc gttccaacgc tggcgcgatc gcttcctgtt tgtggctgat 660
gcaatccaca aatcgcaagc agaaaccggt gaaatcaaag gtcactacct gaacgtgacc 720
gcgccgacct gcgaagaaat gatgaaacgg gctgagttcg ctaaagaact cggcatgccg 780
atcatcatgc atgacttctt gacggctggt ttcaccgcca acaccacctt ggcaaaatgg 840
tgccgcgaca acggcgtcct gctgcacatc caccgtgcaa tgcacgcggt gatcgaccgt 900
cagcgtaacc acgggattca cttccgtgtc ttggccaagt gtttgcgtct gtccggtggt 960
gaccacctcc actccggcac cgtcgtcggc aaactggaag gcgacaaagc ttcgaccttg 1020
ggctttgttg acttgatgcg cgaagaccac atcgaagctg accgcagccg tggggtcttc 1080
ttcacccaag attgggcgtc gatgccgggc gtgctgccgg ttgcttccgg tggtatccac 1140
gtgtggcaca tgcccgcact ggtggaaatc ttcggtgatg actccgttct ccagttcggt 1200
ggcggcacct tgggtcaccc ctggggtaat gctcctggtg caaccgcgaa ccgtgttgcc 1260
ttggaagctt gcgtccaagc tcggaacgaa ggtcgcgacc tctaccgtga aggcggcgac 1320
atccttcgtg aagctggcaa gtggtcgcct gaactggctg ctgccctcga cctctggaaa 1380
gagatcaagt tcgaattcga aacgatggac aagctctaag gagcctctga ctatcgctgg 1440
gggagtgagc gttgctgcgt aaagctttct ccccagcctt tcgacttaac ctttcaggat 1500
ttctgaatca tgagcatgaa aactctgccc aaagagcgtc gtttcgagac tttctcgtac 1560
ctgcctcccc tcagcgatcg ccaaatcgct gcacaaatcg agtacatgat cgagcaaggc 1620
ttccaccct tgatcgagtt caacgagcac tcgaatccgg aagagttcta ctggacgatg 1680
tggaagctcc ccctgtttga ctgcaagagc cctcagcaag tcctcgatga agtgcgtgag 1740
tgccgcagcg aatacggtga ttgctacatc cgtgtcgctg gcttcgacaa catcaagcag 1800
tgccaaaccg tgagcttcat cgttcatcgt cccggccgat actaa 1845

<210> 58
<211> 472
<212> PRT
<213> Artificial

<220>
<223> Com base na sequência traduzida de PCC6301 rbcLS de Synechococcus Sp. do tipo selvagem

<400> 58

Met Ala Lys Thr Gln Ser Ala Ala Gly Tyr Lys Ala Gly Val Lys Pro
1 5 10 15
Tyr Lys Leu Gly Tyr Tyr Thr Pro Asp Tyr Thr Pro Lys Asp Thr Asp
20 25 30
Leu Leu Ala Ala Phe Arg Phe Ser Pro Gln Pro Gly Val Pro Ala Asp
35 40 45
Glu Ala Gly Ala Ala Ile Ala Ala Glu Ser Ser Thr Gly Thr Trp Thr
50 55 60
Thr Val Trp Thr Asp Leu Leu Thr Asp Met Asp Arg Tyr Arg Gly Lys
65 70 75 80
Cys Tyr His Ile Glu Pro Val Gln Gly Glu Glu Asn Ser Tyr Phe Ala

```
                85                  90                  95
Phe Ile Ala Tyr Pro Leu Asp Leu Phe Glu Glu Gly Ser Val Thr Asn
            100                 105                 110
Ile Leu Thr Ser Ile Val Gly Asn Val Phe Gly Phe Lys Ala Ile Arg
        115                 120                 125
Ser Leu Arg Leu Glu Asp Ile Arg Phe Pro Val Ala Leu Val Lys Thr
    130                 135                 140
Phe Gln Gly Pro Pro His Gly Ile Gln Val Glu Arg Asp Leu Leu Asn
145                 150                 155                 160
Lys Tyr Gly Arg Pro Met Leu Gly Cys Thr Ile Lys Pro Lys Leu Gly
                165                 170                 175
Leu Ser Ala Lys Asn Tyr Gly Arg Ala Val Tyr Glu Cys Leu Arg Gly
            180                 185                 190
Gly Leu Asp Phe Thr Lys Asp Asp Glu Asn Ile Asn Ser Gln Pro Phe
        195                 200                 205
Gln Arg Trp Arg Asp Arg Phe Leu Phe Val Ala Asp Ala Ile His Lys
    210                 215                 220
Ser Gln Ala Glu Thr Gly Glu Ile Lys Gly His Tyr Leu Asn Val Thr
225                 230                 235                 240
Ala Pro Thr Cys Glu Glu Met Met Lys Arg Ala Glu Phe Ala Lys Glu
                245                 250                 255
Leu Gly Met Pro Ile Ile Met His Asp Phe Leu Thr Ala Gly Phe Thr
            260                 265                 270
Ala Asn Thr Thr Leu Ala Lys Trp Cys Arg Asp Asn Gly Val Leu Leu
        275                 280                 285
His Ile His Arg Ala Met His Ala Val Ile Asp Arg Gln Arg Asn His
    290                 295                 300
Gly Ile His Phe Arg Val Leu Ala Lys Cys Leu Arg Leu Ser Gly Gly
305                 310                 315                 320
Asp His Leu His Ser Gly Thr Val Val Gly Lys Leu Glu Gly Asp Lys
                325                 330                 335
Ala Ser Thr Leu Gly Phe Val Asp Leu Met Arg Glu Asp His Ile Glu
            340                 345                 350
Ala Asp Arg Ser Arg Gly Val Phe Phe Thr Gln Asp Trp Ala Ser Met
        355                 360                 365
Pro Gly Val Leu Pro Val Ala Ser Gly Gly Ile His Val Trp His Met
    370                 375                 380
Pro Ala Leu Val Glu Ile Phe Gly Asp Asp Ser Val Leu Gln Phe Gly
385                 390                 395                 400
Gly Gly Thr Leu Gly His Pro Trp Gly Asn Ala Pro Gly Ala Thr Ala
                405                 410                 415
Asn Arg Val Ala Leu Glu Ala Cys Val Gln Ala Arg Asn Glu Gly Arg
            420                 425                 430
Asp Leu Tyr Arg Glu Gly Gly Asp Ile Leu Arg Glu Ala Gly Lys Trp
        435                 440                 445
Ser Pro Glu Leu Ala Ala Ala Leu Asp Leu Trp Lys Glu Ile Lys Phe
    450                 455                 460
Glu Phe Glu Thr Met Asp Lys Leu
465                 470
```

<210> 59
<211> 1845
<212> DNA
<213> Artificial

<220>
<223> Com base no nucleotídeo PCC6301 rbcLS de Synechococcus Sp. do tipo selvagem

<400> 59
atggccaaga cgcaatctgc cgcaggctat aaggccgggg tgaaggacta caaactcggc 60

```
tattacaccc ccgattacac ccccaaagac actgacctgc tggcggcttt ccgcttcagc    120
cctcagccgg gtgtccctgc tgacgaagct ggtgcggcga tcgcggctga atcttcgacc    180
ggtacctgga ccaccgtgtg gaccgacttg ctgaccgaca tggatcggta ccgcggcaag    240
tgctaccaca tcgagccggt gcaaggcgaa gagaactcct actttgcgtt catcgcttac    300
ccgctcgacc tgtttgaaga agggtcggtc accaacatcc tgacctcgat catcggtaac    360
gtgtttggct tcaaagctat ccgttcgctg cgtctggaag acatccgctt ccccgtcgcc    420
ttggtcaaaa ccttccaagg tcctccccac ggtatccaag tcgagcgcga cctgctgaac    480
aagtacggcc gtccgatgct gggttgcacg atcaaaccaa aactcggtct gtcggcgaaa    540
aactacggtc gtgccgtcta cgaatgtctg cgcggcggtc tggacttcac caaagacgac    600
gaaaacatca actcgcagcc gttccaacgc tggcgcgatc gcttcctgtt tgtggctgat    660
gcaatccaca aatcgcaagc agaaaccggt gaaatcaaag gtcactacct gaacgtgacc    720
gcgccgacct gcgaagaaat gatgaaacgg gctgagttcg ctaaagaact cggcatgccg    780
atcatcatgc atgacttctt gacggctggt ttcaccgcca acaccacctt ggcaaaatgg    840
tgccgcgaca acggcgtcct gctgcacatc caccgtgcag ggcacgcgac gatcgaccgt    900
cagcgtaacc acgggattca cttccgtgtc ttggccaagt gtttgcgtct gtccggtggt    960
gaccacctcc actccggcac cgtcgtcggc aaactggaag gcgacaaagc ttcgaccttg   1020
ggctttgttg acttgatgcg cgaagaccac atcgaagctg accgcagccg tggggtcttc   1080
ttcacccaag attgggcgtc gatgccgggc gtgctgccgg ttgcttccgg tggtatccac   1140
gtgtggcaca tgcccgcact ggtggaaatc ttcggtgatg actccgttct ccagttcggt   1200
ggcggcacct tgggtcaccc ctggggtaat gctcctggtg caaccgcgaa ccgtgttgcc   1260
ttggaagctt gcgtccaagc tcggaacgaa ggtcgcgacc tctaccgtga aggcggcgac   1320
atccttcgtg aagctggcaa gtggtcgcct gaactggctg ctgccctcga cctctggaaa   1380
gagatcaagt tcgaattcga aacgatggac aagctctaag gagcctctga ctatcgctgg   1440
gggagtgagc gttgctgcgt aaagctttct ccccagcctt tcgacttaac ctttcaggat   1500
ttctgaatca tgagcatgaa aactctgccc aaagagcgtc gtttcgagac tttctcgtac   1560
ctgcctcccc tcagcgatcg ccaaatcgct gcacaaatcg agtacatgat cgagcaaggc   1620
ttccaccccc tgatcgagtt caacgagcac tcgaatccgg aagagttcta ctggacgatg   1680
tggaagctcc ccctgtttga ctgcaagagc cctcagcaag tcctcgatga agtgcgtgag   1740
tgccgcagcg aatacggtga ttgctacatc cgtgtcgctg gcttcgacaa catcaagcag   1800
tgccaaaccg tgagcttcat cgttcatcgt cccggccgat actaa                   1845
```

<210> 60
<211> 472
<212> PRT
<213> Artificial

<220>
<223> Com base na sequência traduzida de PCC6301 rbcLS de Synechococcus Sp. do tipo selvagem

<400> 60

```
Met Ala Lys Thr Gln Ser Ala Ala Gly Tyr Lys Ala Gly Val Lys Asp
1               5                  10                  15
Tyr Lys Leu Gly Tyr Tyr Thr Pro Asp Tyr Thr Pro Lys Asp Thr Asp
            20                  25                  30
Leu Leu Ala Ala Phe Arg Phe Ser Pro Gln Pro Gly Val Pro Ala Asp
        35                  40                  45
Glu Ala Gly Ala Ala Ile Ala Ala Glu Ser Ser Thr Gly Thr Trp Thr
    50                  55                  60
Thr Val Trp Thr Asp Leu Leu Thr Asp Met Asp Arg Tyr Arg Gly Lys
65                  70                  75                  80
Cys Tyr His Ile Glu Pro Val Gln Gly Glu Glu Asn Ser Tyr Phe Ala
                85                  90                  95
Phe Ile Ala Tyr Pro Leu Asp Leu Phe Glu Glu Gly Ser Val Thr Asn
            100                 105                 110
Ile Leu Thr Ser Ile Ile Gly Asn Val Phe Gly Phe Lys Ala Ile Arg
        115                 120                 125
Ser Leu Arg Leu Glu Asp Ile Arg Phe Pro Val Ala Leu Val Lys Thr
    130                 135                 140
Phe Gln Gly Pro Pro His Gly Ile Gln Val Glu Arg Asp Leu Leu Asn
```

```
145                 150                 155                 160
Lys Tyr Gly Arg Pro Met Leu Gly Cys Thr Ile Lys Pro Lys Leu Gly
                165                 170                 175
Leu Ser Ala Lys Asn Tyr Gly Arg Ala Val Tyr Glu Cys Leu Arg Gly
            180                 185                 190
Gly Leu Asp Phe Thr Lys Asp Asp Glu Asn Ile Asn Ser Gln Pro Phe
        195                 200                 205
Gln Arg Trp Arg Asp Arg Phe Leu Phe Val Ala Asp Ala Ile His Lys
    210                 215                 220
Ser Gln Ala Glu Thr Gly Glu Ile Lys Gly His Tyr Leu Asn Val Thr
225                 230                 235                 240
Ala Pro Thr Cys Glu Glu Met Met Lys Arg Ala Glu Phe Ala Lys Glu
                245                 250                 255
Leu Gly Met Pro Ile Ile Met His Asp Phe Leu Thr Ala Gly Phe Thr
            260                 265                 270
Ala Asn Thr Thr Leu Ala Lys Trp Cys Arg Asp Asn Gly Val Leu Leu
        275                 280                 285
His Ile His Arg Ala Gly His Ala Thr Ile Asp Arg Gln Arg Asn His
    290                 295                 300
Gly Ile His Phe Arg Val Leu Ala Lys Cys Leu Arg Leu Ser Gly Gly
305                 310                 315                 320
Asp His Leu His Ser Gly Thr Val Val Gly Lys Leu Glu Gly Asp Lys
                325                 330                 335
Ala Ser Thr Leu Gly Phe Val Asp Leu Met Arg Glu Asp His Ile Glu
            340                 345                 350
Ala Asp Arg Ser Arg Gly Val Phe Phe Thr Gln Asp Trp Ala Ser Met
        355                 360                 365
Pro Gly Val Leu Pro Val Ala Ser Gly Gly Ile His Val Trp His Met
    370                 375                 380
Pro Ala Leu Val Glu Ile Phe Gly Asp Asp Ser Val Leu Gln Phe Gly
385                 390                 395                 400
Gly Gly Thr Leu Gly His Pro Trp Gly Asn Ala Pro Gly Ala Thr Ala
                405                 410                 415
Asn Arg Val Ala Leu Glu Ala Cys Val Gln Ala Arg Asn Glu Gly Arg
            420                 425                 430
Asp Leu Tyr Arg Glu Gly Gly Asp Ile Leu Arg Glu Ala Gly Lys Trp
        435                 440                 445
Ser Pro Glu Leu Ala Ala Ala Leu Asp Leu Trp Lys Glu Ile Lys Phe
    450                 455                 460
Glu Phe Glu Thr Met Asp Lys Leu
465                 470
```

<210> 61
<211> 1845
<212> DNA
<213> Artificial

<220>
<223> Com base no nucleotídeo PCC6301 rbcLS de Synechococcus Sp. do tipo selvagem

<400> 61

```
atggccaaga cgcaatctgc cgcaggctat aaggccgggg tgaaggacta caaactcggc     60
tattacaccc ccgattacac ccccaaagac actgacctgc tggcggcttt ccgcttcagc    120
cctcagccgg gtgtccctgc tgacgaagct ggtgcggcga tcgcggctga atcttcgacc    180
ggtacctgga ccaccgtgtg gaccgacttg ctgaccgaca tggatcggta caaaggcaag    240
tgctaccaca tcgagccggt gcaaggcgaa gagaactcct actttgcgtt catcgcttac    300
ccgctcgacc tgtttgaaga agggtcggtc accaacatcc tgacctcgat cgtcggtaac    360
gtgtttggct tcaaagctat ccgttcgctg cgtctggaag acatccgctt ccccgtcgcc    420
ttggtcaaaa ccttccaagg tcctccccac ggtatccaag tcgagcgcga cctgctgaac    480
```

```
aagtacggcc gtccgatgct gggttgcacg atcaaaccaa aactcggtct gtcggcgaaa    540
aactacggtc gtgccgtcta cgaatgtctg cgcggcggtc tggacttcac caaagacgac    600
gaaaacatca actcgcagcc gttccaacgc tggcgcgatc gcttcctgtt tgtggctgat    660
gcaatccaca aatcgcaagc agaaaccggt gaaatcaaag gtcactacct gaacgtgacc    720
gcgccgacct gcgaagaaat gatgaaacgg gctgagttcg ctaaagaact cggcatgccg    780
atcatcatgc atgacttctt gacggctggt ttcaccgcca acaccacctt ggcaaaatgg    840
tgccgcgaca acggcgtcct gctgcacatc caccgtgcaa tgcacgcggt gatcgaccgt    900
cagcgtaacc acgggattca cttccgtgtc ttggccaagt gtttgcgtct gtccggtggt    960
gaccacctcc actccggcac cgtcgtcggc aaactggaag gcgacaaagc ttcgaccttg   1020
ggctttgttg acttgatgcg cgaagaccac atcgaagctg accgcagccg tggggtcttc   1080
ttcacccaag attgggcgtc gatgccgggc gtgctgccgg ttgcttccgg tggtatccac   1140
gtgtggcaca tgcccgcact ggtggaaatc ttcggtgatg actccgttct ccagttcggt   1200
ggcggcacct tgggtcaccc ctggggtaat gctcctggtg caaccgcgaa ccgtgttgcc   1260
ttggaagctt gcgtccaagc tcggaacgaa ggtcgcgacc tctaccgtga aggcggcgac   1320
atccttcgtg aagctggcaa gtggtcgcct gaactggctg ctgccctcga cctctggaaa   1380
gagatcaagt tcgaattcga aacgatggac aagctctaag gagcctctga ctatcgctgg   1440
gggagtgagc gttgctgcgt aaagctttct ccccagcctt tcgacttaac ctttcaggat   1500
ttctgaatca tgagcatgaa aactctgccc aaagagcgtc gtttcgagac tttctcgtac   1560
ctgcctcccc tcagcgatcg ccaaatcgct gcacaaatcg agtacatgat cgagcaaggc   1620
ttccaccccc tgatcgagtt caacgagcac tcgaatccgg aagagttcta ctggacgatg   1680
tggaagctcc ccctgtttga ctgcaagagc cctcagcaag tcctcgatga agtgcgtgag   1740
tgccgcagcg aatacggtga ttgctacatc cgtgtcgctg gcttcgacaa catcaagcag   1800
tgccaaaccg tgagcttcat cgttcatcgt cccggccgat actaa                   1845
```

<210> 62
<211> 472
<212> PRT
<213> Artificial

<220>
<223> Com base na sequência traduzida de PCC6301 rbcLS de Synechococcus Sp. do tipo selvagem

<400> 62

```
Met Ala Lys Thr Gln Ser Ala Ala Gly Tyr Lys Ala Gly Val Lys Asp
1               5                  10                  15
Tyr Lys Leu Gly Tyr Tyr Thr Pro Asp Tyr Thr Pro Lys Asp Thr Asp
            20                  25                  30
Leu Leu Ala Ala Phe Arg Phe Ser Pro Gln Pro Gly Val Pro Ala Asp
        35                  40                  45
Glu Ala Gly Ala Ala Ile Ala Ala Glu Ser Ser Thr Gly Thr Trp Thr
    50                  55                  60
Thr Val Trp Thr Asp Leu Leu Thr Asp Met Asp Arg Tyr Lys Gly Lys
65                  70                  75                  80
Cys Tyr His Ile Glu Pro Val Gln Gly Glu Glu Asn Ser Tyr Phe Ala
                85                  90                  95
Phe Ile Ala Tyr Pro Leu Asp Leu Phe Glu Glu Gly Ser Val Thr Asn
            100                 105                 110
Ile Leu Thr Ser Ile Val Gly Asn Val Phe Gly Phe Lys Ala Ile Arg
        115                 120                 125
Ser Leu Arg Leu Glu Asp Ile Arg Phe Pro Val Ala Leu Val Lys Thr
    130                 135                 140
Phe Gln Gly Pro Pro His Gly Ile Gln Val Glu Arg Asp Leu Leu Asn
145                 150                 155                 160
Lys Tyr Gly Arg Pro Met Leu Gly Cys Thr Ile Lys Pro Lys Leu Gly
                165                 170                 175
Leu Ser Ala Lys Asn Tyr Gly Arg Ala Val Tyr Glu Cys Leu Arg Gly
            180                 185                 190
Gly Leu Asp Phe Thr Lys Asp Asp Glu Asn Ile Asn Ser Gln Pro Phe
        195                 200                 205
```

```
Gln Arg Trp Arg Asp Arg Phe Leu Phe Val Ala Asp Ala Ile His Lys
    210                 215                 220
Ser Gln Ala Glu Thr Gly Glu Ile Lys Gly His Tyr Leu Asn Val Thr
225                 230                 235                 240
Ala Pro Thr Cys Glu Glu Met Met Lys Arg Ala Glu Phe Ala Lys Glu
                245                 250                 255
Leu Gly Met Pro Ile Ile Met His Asp Phe Leu Thr Ala Gly Phe Thr
            260                 265                 270
Ala Asn Thr Thr Leu Ala Lys Trp Cys Arg Asp Asn Gly Val Leu Leu
        275                 280                 285
His Ile His Arg Ala Met His Ala Val Ile Asp Arg Gln Arg Asn His
    290                 295                 300
Gly Ile His Phe Arg Val Leu Ala Lys Cys Leu Arg Leu Ser Gly Gly
305                 310                 315                 320
Asp His Leu His Ser Gly Thr Val Val Gly Lys Leu Glu Gly Asp Lys
                325                 330                 335
Ala Ser Thr Leu Gly Phe Val Asp Leu Met Arg Glu Asp His Ile Glu
            340                 345                 350
Ala Asp Arg Ser Arg Gly Val Phe Phe Thr Gln Asp Trp Ala Ser Met
        355                 360                 365
Pro Gly Val Leu Pro Val Ala Ser Gly Gly Ile His Val Trp His Met
    370                 375                 380
Pro Ala Leu Val Glu Ile Phe Gly Asp Asp Ser Val Leu Gln Phe Gly
385                 390                 395                 400
Gly Gly Thr Leu Gly His Pro Trp Gly Asn Ala Pro Gly Ala Thr Ala
                405                 410                 415
Asn Arg Val Ala Leu Glu Ala Cys Val Gln Ala Arg Asn Glu Gly Arg
            420                 425                 430
Asp Leu Tyr Arg Glu Gly Gly Asp Ile Leu Arg Glu Ala Gly Lys Trp
        435                 440                 445
Ser Pro Glu Leu Ala Ala Ala Leu Asp Leu Trp Lys Glu Ile Lys Phe
    450                 455                 460
Glu Phe Glu Thr Met Asp Lys Leu
465                 470
```

<210> 63
<211> 1845
<212> DNA
<213> Artificial

<220>
<223> Com base no nucleotídeo PCC6301 rbcLS de Synechococcus Sp. do tipo selvagem

<400> 63

```
atggccaaga cgcaatctgc cgcaggctat aaggccgggg tgaaggacta caaactcacc     60
tattacaccc ccgattacac ccccaaagac actgacctgc tggcggcttt ccgcttcagc    120
cctcagccgg gtgtccctgc tgacgaagct ggtgcggcga tcgcggctga atcttcgacc    180
ggtacctgga ccaccgtgtg gaccgacttg ctgaccgaca tggatcggta ccgcggcaag    240
tgctaccaca tcgagccggt gcaaggcgaa gagaactcct actttgcgtt catcgcttac    300
ccgctcgacc tgtttgaaga agggtcggtc accaacatcc tgacctcgat cgtcggtaac    360
gtgtttggct tcaaagctat ccgttcgctg cgtctggaag acatccgctt ccccgtcgcc    420
ttggtcaaaa ccttccaagg tcctccccac ggtatccaag tcgagcgcga cctgctgaac    480
aagtacggcc gtccgatgct gggttgcacg atcaaaccaa aactcggtct gtcggcgaaa    540
aactacggtc gtgccgtcta cgaatgtctg cgcggcggtc tggacttcac caaagacgac    600
gaaaacatca actcgcagcc gttccaacgc tggcgcgatc gcttcctgtt tgtggctgat    660
gcaatccaca aatcgcaagc agaaaccggt gaaatcaaag gtcactacct gaacgtgacc    720
gcgccgacct gcgaagaaat gatgaaacgg gctgagttcg ctaaagaact cggcatgccg    780
atcatcatgc atgacttctt gacggctggt ttcaccgcca acaccacctt ggcaaaatgg    840
tgccgcgaca acggcgtcct gctgcacatc caccgtgcaa tgcacgcggt gatcgaccgt    900
cagcgtaacc acgggattca cttccgtgtc ttggccaagt gtttgcgtct gtccggtggt    960
```

```
gaccacctcc actccggcac cgtcgtcggc aaactggaag gcgacaaagc ttcgaccttg   1020
ggctttgttg acttgatgcg cgaagaccac atcgaagctg accgcagccg tggggtcttc   1080
ttcacccaag attgggcgtc gatgccgggc gtgctgccgg ttgcttccgg tggtatccac   1140
gtgtggcaca tgcccgcact ggtggaaatc ttcggtgatg actccgttct ccagttcggt   1200
ggcggcacct tgggtcaccc ctggggtaat gctcctggtg caaccgcgaa ccgtgttgcc   1260
ttggaagctt gcgtccaagc tcggaacgaa ggtcgcgacc tctaccgtga aggcggcgac   1320
atccttcgtg aagctggcaa gtggtcgcct gaactggctg ctgccctcga cctctggaaa   1380
gagatcaagt tcgaattcga aacgatggac aagctctaag gagcctctga ctatcgctgg   1440
gggagtgagc gttgctgcgt aaagctttct ccccagcctt tcgacttaac ctttcaggat   1500
ttctgaatca tgagcatgaa aactctgccc aaagagcgtc gtttcgagac tttctcgtac   1560
ctgcctcccc tcagcgatcg ccaaatcgct gcacaaatcg agtacatgat cgagcaaggc   1620
ttccaccccc tgatcgagtt caacgagcac tcgaatccgg aagagttcta ctggacgatg   1680
tggaagctcc ccctgtttga ctgcaagagc cctcagcaag tcctcgatga agtgcgtgag   1740
tgccgcagcg aatacggtga ttgctacatc cgtgtcgctg gcttcgacaa catcaagcag   1800
tgccaaaccg tgagcttcat cgttcatcgt cccggccgat actaa                   1845
```

<210> 64
<211> 472
<212> PRT
<213> Artificial

<220>
<223> Com base na sequência traduzida de PCC6301 rbcLS de Synechococcus Sp. do tipo selvagem

<400> 64

```
Met Ala Lys Thr Gln Ser Ala Ala Gly Tyr Lys Ala Gly Val Lys Asp
1               5                  10                  15
Tyr Lys Leu Thr Tyr Tyr Thr Pro Asp Tyr Thr Pro Lys Asp Thr Asp
            20                  25                  30
Leu Leu Ala Ala Phe Arg Phe Ser Pro Gln Pro Gly Val Pro Ala Asp
        35                  40                  45
Glu Ala Gly Ala Ala Ile Ala Ala Glu Ser Ser Thr Gly Thr Trp Thr
    50                  55                  60
Thr Val Trp Thr Asp Leu Leu Thr Asp Met Asp Arg Tyr Arg Gly Lys
65                  70                  75                  80
Cys Tyr His Ile Glu Pro Val Gln Gly Glu Glu Asn Ser Tyr Phe Ala
                85                  90                  95
Phe Ile Ala Tyr Pro Leu Asp Leu Phe Glu Glu Gly Ser Val Thr Asn
            100                 105                 110
Ile Leu Thr Ser Ile Val Gly Asn Val Phe Gly Phe Lys Ala Ile Arg
        115                 120                 125
Ser Leu Arg Leu Glu Asp Ile Arg Phe Pro Val Ala Leu Val Lys Thr
    130                 135                 140
Phe Gln Gly Pro Pro His Gly Ile Gln Val Glu Arg Asp Leu Leu Asn
145                 150                 155                 160
Lys Tyr Gly Arg Pro Met Leu Gly Cys Thr Ile Lys Pro Lys Leu Gly
                165                 170                 175
Leu Ser Ala Lys Asn Tyr Gly Arg Ala Val Tyr Glu Cys Leu Arg Gly
            180                 185                 190
Gly Leu Asp Phe Thr Lys Asp Asp Glu Asn Ile Asn Ser Gln Pro Phe
        195                 200                 205
Gln Arg Trp Arg Asp Arg Phe Leu Phe Val Ala Asp Ala Ile His Lys
    210                 215                 220
Ser Gln Ala Glu Thr Gly Glu Ile Lys Gly His Tyr Leu Asn Val Thr
225                 230                 235                 240
Ala Pro Thr Cys Glu Glu Met Met Lys Arg Ala Glu Phe Ala Lys Glu
                245                 250                 255
Leu Gly Met Pro Ile Ile Met His Asp Phe Leu Thr Ala Gly Phe Thr
            260                 265                 270
```

```
Ala Asn Thr Thr Leu Ala Lys Trp Cys Arg Asp Asn Gly Val Leu Leu
        275                 280                 285
His Ile His Arg Ala Met His Ala Val Ile Asp Arg Gln Arg Asn His
    290                 295                 300
Gly Ile His Phe Arg Val Leu Ala Lys Cys Leu Arg Leu Ser Gly Gly
305                 310                 315                 320
Asp His Leu His Ser Gly Thr Val Val Gly Lys Leu Glu Gly Asp Lys
                325                 330                 335
Ala Ser Thr Leu Gly Phe Val Asp Leu Met Arg Glu Asp His Ile Glu
            340                 345                 350
Ala Asp Arg Ser Arg Gly Val Phe Phe Thr Gln Asp Trp Ala Ser Met
        355                 360                 365
Pro Gly Val Leu Pro Val Ala Ser Gly Gly Ile His Val Trp His Met
    370                 375                 380
Pro Ala Leu Val Glu Ile Phe Gly Asp Asp Ser Val Leu Gln Phe Gly
385                 390                 395                 400
Gly Gly Thr Leu Gly His Pro Trp Gly Asn Ala Pro Gly Ala Thr Ala
                405                 410                 415
Asn Arg Val Ala Leu Glu Ala Cys Val Gln Ala Arg Asn Glu Gly Arg
            420                 425                 430
Asp Leu Tyr Arg Glu Gly Gly Asp Ile Leu Arg Glu Ala Gly Lys Trp
        435                 440                 445
Ser Pro Glu Leu Ala Ala Ala Leu Asp Leu Trp Lys Glu Ile Lys Phe
    450                 455                 460
Glu Phe Glu Thr Met Asp Lys Leu
465                 470
```

<210> 65
<211> 1845
<212> DNA
<213> Artificial

<220>
<223> Com base no nucleotídeo PCC6301 rbcLS de Synechococcus Sp. do tipo selvagem

<400> 65

```
atggccaaga cgcaatctgc cgcaggctat aaggccgggg tgaaggacta caaactcacc     60
tattacaccc ccgattacac ccccaaagac actgacctgc tggcggcttt ccgcttcagc    120
cctcagccgg gtgtccctgc tgacgaagct ggtgcggcga tcgcggctga atcttcgacc    180
ggtacctgga ccaccgtgtg gaccgacttg ctgaccgaca tggatcggta caaaggcaag    240
tgctaccaca tcgagccggt gcaaggcgaa gagaactcct actttgcgtt catcgcttac    300
ccgctcgacc tgtttgaaga agggtcggtc accaacatcc tgacctcgat catcggtaac    360
gtgtttggct tcaaagctat ccgttcgctg cgtctggaag acatccgctt ccccgtcgcc    420
ttggtcaaaa ccttccaagg tcctccccac ggtatccaag tcgagcgcga cctgctgaac    480
aagtacggcc gtccgatgct gggttgcacg atcaaaccaa aactcggtct gtcggcgaaa    540
aactacggtc gtgccgtcta cgaatgtctg cgcggcggtc tggacttcac caaagacgac    600
gaaaacatca actcgcagcc gttccaacgc tggcgcgatc gcttcctgtt tgtggctgat    660
gcaatccaca aatcgcaagc agaaaccggt gaaatcaaag gtcactacct gaacgtgacc    720
gcgccgacct gcgaagaaat gatgaaacgg gctgagttcg ctaaagaact cggcatgccg    780
atcatcatgc atgacttctt gacggctggt ttcaccgcca acaccacctt ggcaaaatgg    840
tgccgcgaca acggcgtcct gctgcacatc caccgtgcag ggcacgcggt gatcgaccgt    900
cagcgtaacc acgggattca cttccgtgtc ttggccaagt gtttgcgtct gtccggtggt    960
gaccacctcc actccggcac cgtcgtcggc aaactggaag gcgacaaagc ttcgaccttg   1020
ggctttgttg acttgatgcg cgaagaccac atcgaagctg accgcagccg tggggtcttc   1080
ttcacccaag attgggcgtc gatgccgggc gtgctgccgg ttgcttccgg tggtatccac   1140
gtgtggcaca tgcccgcact ggtggaaatc ttcggtgatg actccgttct ccagttcggt   1200
ggcggcacct tgggtcaccc ctggggtaat gctcctggtg caaccgcgaa ccgtgttgcc   1260
ttggaagctt gcgtccaagc tcggaacgaa ggtcgcgacc tctaccgtga aggcggcgac   1320
atccttcgtg aagctggcaa gtggtcgcct gaactggctg ctgccctcga cctctggaaa   1380
gagatcaagt tcgaattcga aacgatggac aagctctaag gagcctctga ctatcgctgg   1440
```

```
gggagtgagc gttgctgcgt aaagctttct ccccagcctt tcgacttaac ctttcaggat   1500
ttctgaatca tgagcatgaa aactctgccc aaagagcgtc gtttcgagac tttctcgtac   1560
ctgcctcccc tcagcgatcg ccaaatcgct gcacaaatcg agtacatgat cgagcaaggc   1620
ttccacccct tgatcgagtt caacgagcac tcgaatccgg aagagttcta ctggacgatg   1680
tggaagctcc ccctgtttga ctgcaagagc cctcagcaag tcctcgatga agtgcgtgag   1740
tgccgcagcg aatacggtga ttgctacatc cgtgtcgctg gcttcgacaa catcaagcag   1800
tgccaaaccg tgagcttcat cgttcatcgt cccggccgat actaa                   1845
```

<210> 66
<211> 472
<212> PRT
<213> Artificial

<220>
<223> Com base na sequência traduzida de PCC6301 rbcLS de Synechococcus Sp. do tipo selvagem

<400> 66

```
Met Ala Lys Thr Gln Ser Ala Ala Gly Tyr Lys Ala Gly Val Lys Asp
1               5                  10                  15
Tyr Lys Leu Thr Tyr Tyr Thr Pro Asp Tyr Thr Pro Lys Asp Thr Asp
            20                  25                  30
Leu Leu Ala Ala Phe Arg Phe Ser Pro Gln Pro Gly Val Pro Ala Asp
        35                  40                  45
Glu Ala Gly Ala Ala Ile Ala Ala Glu Ser Ser Thr Gly Thr Trp Thr
    50                  55                  60
Thr Val Trp Thr Asp Leu Leu Thr Asp Met Asp Arg Tyr Lys Gly Lys
65                  70                  75                  80
Cys Tyr His Ile Glu Pro Val Gln Gly Glu Glu Asn Ser Tyr Phe Ala
                85                  90                  95
Phe Ile Ala Tyr Pro Leu Asp Leu Phe Glu Glu Gly Ser Val Thr Asn
            100                 105                 110
Ile Leu Thr Ser Ile Ile Gly Asn Val Phe Gly Phe Lys Ala Ile Arg
        115                 120                 125
Ser Leu Arg Leu Glu Asp Ile Arg Phe Pro Val Ala Leu Val Lys Thr
    130                 135                 140
Phe Gln Gly Pro Pro His Gly Ile Gln Val Glu Arg Asp Leu Leu Asn
145                 150                 155                 160
Lys Tyr Gly Arg Pro Met Leu Gly Cys Thr Ile Lys Pro Lys Leu Gly
                165                 170                 175
Leu Ser Ala Lys Asn Tyr Gly Arg Ala Val Tyr Glu Cys Leu Arg Gly
            180                 185                 190
Gly Leu Asp Phe Thr Lys Asp Asp Glu Asn Ile Asn Ser Gln Pro Phe
        195                 200                 205
Gln Arg Trp Arg Asp Arg Phe Leu Phe Val Ala Asp Ala Ile His Lys
    210                 215                 220
Ser Gln Ala Glu Thr Gly Glu Ile Lys Gly His Tyr Leu Asn Val Thr
225                 230                 235                 240
Ala Pro Thr Cys Glu Glu Met Met Lys Arg Ala Glu Phe Ala Lys Glu
                245                 250                 255
Leu Gly Met Pro Ile Ile Met His Asp Phe Leu Thr Ala Gly Phe Thr
            260                 265                 270
Ala Asn Thr Thr Leu Ala Lys Trp Cys Arg Asp Asn Gly Val Leu Leu
        275                 280                 285
His Ile His Arg Ala Gly His Ala Val Ile Asp Arg Gln Arg Asn His
    290                 295                 300
Gly Ile His Phe Arg Val Leu Ala Lys Cys Leu Arg Leu Ser Gly Gly
305                 310                 315                 320
Asp His Leu His Ser Gly Thr Val Val Gly Lys Leu Glu Gly Asp Lys
                325                 330                 335
```

```
Ala Ser Thr Leu Gly Phe Val Asp Leu Met Arg Glu Asp His Ile Glu
            340                 345                 350
Ala Asp Arg Ser Arg Gly Val Phe Phe Thr Gln Asp Trp Ala Ser Met
        355                 360                 365
Pro Gly Val Leu Pro Val Ala Ser Gly Gly Ile His Val Trp His Met
    370                 375                 380
Pro Ala Leu Val Glu Ile Phe Gly Asp Asp Ser Val Leu Gln Phe Gly
385                 390                 395                 400
Gly Gly Thr Leu Gly His Pro Trp Gly Asn Ala Pro Gly Ala Thr Ala
                405                 410                 415
Asn Arg Val Ala Leu Glu Ala Cys Val Gln Ala Arg Asn Glu Gly Arg
            420                 425                 430
Asp Leu Tyr Arg Glu Gly Gly Asp Ile Leu Arg Glu Ala Gly Lys Trp
        435                 440                 445
Ser Pro Glu Leu Ala Ala Ala Leu Asp Leu Trp Lys Glu Ile Lys Phe
    450                 455                 460
Glu Phe Glu Thr Met Asp Lys Leu
465                 470
```

<210> 67
<211> 1845
<212> DNA
<213> Artificial

<220>
<223> Com base no nucleotídeo PCC6301 rbcLS de Synechococcus Sp. do tipo selvagem

<400> 67

```
atggccaaga cgcaatctgc cgcaggctat aaggccgggg tgaaggacta caaactcacc     60
tattacaccc ccgattacac ccccaaagac actgacctgc tggcggcttt ccgcttcagc    120
cctcagccgg gtgtccctgc tgacgaagct ggtgcggcga tcgcggctga atcttcgacc    180
ggtacctgga ccaccgtgtg gaccgacttg ctgaccgaca tggatcggta caaaggcaag    240
tgctaccaca tcgagccggt gcaaggcgaa gagaactcct actttgcgtt catcgcttac    300
ccgctcgacc tgtttgaaga agggtcggtc accaacatcc tgacctcgat cgtcggtaac    360
gtgtttggct tcaaagctat ccgttcgctg cgtctggaag acatccgctt ccccgtcgcc    420
ttggtcaaaa ccttccaagg tcctccccac ggtatccaag tcgagcgcga cctgctgaac    480
aagtacggcc gtccgatgct gggttgcacg atcaaaccaa aactcggtct gtcggcgaaa    540
aactacggtc gtgccgtcta cgaatgtctg cgcggcggtc tggacttcac caaagacgac    600
gaaaacatca actcgcagcc gttccaacgc tggcgcgatc gcttcctgtt tgtggctgat    660
gcaatccaca aatcgcaagc agaaaccggt gaaatcaaag gtcactacct gaacgtgacc    720
gcgccgacct gcgaagaaat gatgaaacgg gctgagttcg ctaaagaact cggcatgccg    780
atcatcatgc atgacttctt gacggctggt ttcaccgcca acaccacctt ggcaaaatgg    840
tgccgcgaca acggcgtcct gctgcacatc caccgtgcag ggcacgcggt gatcgaccgt    900
cagcgtaacc acgggattca cttccgtgtc ttggccaagt gtttgcgtct gtccggtggt    960
gaccacctcc actccggcac cgtcgtcggc aaactggaag gcgacaaagc ttcgaccttg   1020
ggctttgttg acttgatgcg cgaagaccac atcgaagctg accgcagccg tggggtcttc   1080
ttcacccaag attgggcgtc gatgccgggc gtgctgccgg ttgcttccgg tggtatccac   1140
gtgtggcaca tgcccgcact ggtggaaatc ttcggtgatg actccgttct ccagttcggt   1200
ggcggcacct tgggtcaccc ctggggtaat gctcctggtg caaccgcgaa ccgtgttgcc   1260
ttggaagctt gcgtccaagc tcggaacgaa ggtcgcgacc tctaccgtga aggcggcgac   1320
atccttcgtg aagctggcaa gtggtcgcct gaactggctg ctgccctcga cctctggaaa   1380
gagatcaagt tcgaattcga aacgatggac aagctctaag gagcctctga ctatcgctgg   1440
gggagtgagc gttgctgcgt aaagctttct ccccagcctt tcgacttaac ctttcaggat   1500
ttctgaatca tgagcatgaa aactctgccc aaagagcgtc gtttcgagac tttctcgtac   1560
ctgcctcccc tcagcgatcg ccaaatcgct gcacaaatcg agtacatgat cgagcaaggc   1620
ttccaccccc tgatcgagtt caacgagcac tcgaatccgg aagagttcta ctggacgatg   1680
tggaagctcc ccctgtttga ctgcaagagc cctcagcaag tcctcgatga agtgcgtgag   1740
tgccgcagcg aatacggtga ttgctacatc cgtgtcgctg gcttcgacaa catcaagcag   1800
tgccaaaccg tgagcttcat cgttcatcgt cccggccgat actaa                   1845
```

<210> 68
<211> 472
<212> PRT
<213> Artificial

<220>
<223> Com base na sequência traduzida de PCC6301 rbcLS de Synechococcus Sp. do tipo selvagem

<400> 68

```
Met Ala Lys Thr Gln Ser Ala Ala Gly Tyr Lys Ala Gly Val Lys Asp
1               5                  10                  15
Tyr Lys Leu Thr Tyr Tyr Thr Pro Asp Tyr Thr Pro Lys Asp Thr Asp
            20                  25                  30
Leu Leu Ala Ala Phe Arg Phe Ser Pro Gln Pro Gly Val Pro Ala Asp
        35                  40                  45
Glu Ala Gly Ala Ala Ile Ala Ala Glu Ser Ser Thr Gly Thr Trp Thr
    50                  55                  60
Thr Val Trp Thr Asp Leu Leu Thr Asp Met Asp Arg Tyr Lys Gly Lys
65                  70                  75                  80
Cys Tyr His Ile Glu Pro Val Gln Gly Glu Glu Asn Ser Tyr Phe Ala
                85                  90                  95
Phe Ile Ala Tyr Pro Leu Asp Leu Phe Glu Glu Gly Ser Val Thr Asn
            100                 105                 110
Ile Leu Thr Ser Ile Val Gly Asn Val Phe Gly Phe Lys Ala Ile Arg
        115                 120                 125
Ser Leu Arg Leu Glu Asp Ile Arg Phe Pro Val Ala Leu Val Lys Thr
    130                 135                 140
Phe Gln Gly Pro Pro His Gly Ile Gln Val Glu Arg Asp Leu Leu Asn
145                 150                 155                 160
Lys Tyr Gly Arg Pro Met Leu Gly Cys Thr Ile Lys Pro Lys Leu Gly
                165                 170                 175
Leu Ser Ala Lys Asn Tyr Gly Arg Ala Val Tyr Glu Cys Leu Arg Gly
            180                 185                 190
Gly Leu Asp Phe Thr Lys Asp Asp Glu Asn Ile Asn Ser Gln Pro Phe
        195                 200                 205
Gln Arg Trp Arg Asp Arg Phe Leu Phe Val Ala Asp Ala Ile His Lys
    210                 215                 220
Ser Gln Ala Glu Thr Gly Glu Ile Lys Gly His Tyr Leu Asn Val Thr
225                 230                 235                 240
Ala Pro Thr Cys Glu Glu Met Met Lys Arg Ala Glu Phe Ala Lys Glu
                245                 250                 255
Leu Gly Met Pro Ile Ile Met His Asp Phe Leu Thr Ala Gly Phe Thr
            260                 265                 270
Ala Asn Thr Thr Leu Ala Lys Trp Cys Arg Asp Asn Gly Val Leu Leu
        275                 280                 285
His Ile His Arg Ala Gly His Ala Val Ile Asp Arg Gln Arg Asn His
    290                 295                 300
Gly Ile His Phe Arg Val Leu Ala Lys Cys Leu Arg Leu Ser Gly Gly
305                 310                 315                 320
Asp His Leu His Ser Gly Thr Val Val Gly Lys Leu Glu Gly Asp Lys
                325                 330                 335
Ala Ser Thr Leu Gly Phe Val Asp Leu Met Arg Glu Asp His Ile Glu
            340                 345                 350
Ala Asp Arg Ser Arg Gly Val Phe Phe Thr Gln Asp Trp Ala Ser Met
        355                 360                 365
Pro Gly Val Leu Pro Val Ala Ser Gly Gly Ile His Val Trp His Met
    370                 375                 380
Pro Ala Leu Val Glu Ile Phe Gly Asp Asp Ser Val Leu Gln Phe Gly
385                 390                 395                 400
```

```
Gly Gly Thr Leu Gly His Pro Trp Gly Asn Ala Pro Gly Ala Thr Ala
                405                 410                 415
Asn Arg Val Ala Leu Glu Ala Cys Val Gln Ala Arg Asn Glu Gly Arg
            420                 425                 430
Asp Leu Tyr Arg Glu Gly Gly Asp Ile Leu Arg Glu Ala Gly Lys Trp
        435                 440                 445
Ser Pro Glu Leu Ala Ala Ala Leu Asp Leu Trp Lys Glu Ile Lys Phe
    450                 455                 460
Glu Phe Glu Thr Met Asp Lys Leu
465                 470
```

<210> 69
<211> 1845
<212> DNA
<213> Artificial

<220>
<223> Com base no nucleotídeo PCC6301 rbcLS de Synechococcus Sp. do tipo selvagem

<400> 69

```
atggccaaga cgcaatctgc cgcaggctat aaggccgggg tgaaggacta caaactcacc     60
tattacaccc ccgattacac ccccaaagac actgacctgc tggcggcttt ccgcttcagc    120
cctcagccgg gtgtccctgc tgacgaagct ggtgcggcga tcgcggctga atcttcgacc    180
ggtacctgga ccaccgtgtg gaccgacttg ctgaccgaca tggatcggta caaaggcaag    240
tgctaccaca tcgagccggt gcaaggcgaa gagaactcct actttgcgtt catcgcttac    300
ccgctcgacc tgtttgaaga agggtcggtc accaacatcc tgacctcgat catcggtaac    360
gtgtttggct tcaaagctat ccgttcgctg cgtctggaag acatccgctt ccccgtcgcc    420
ttggtcaaaa ccttccaagg tcctccccac ggtatccaag tcgagcgcga cctgctgaac    480
aagtacggcc gtccgatgct gggttgcacg atcaaaccaa aactcggtct gtcggcgaaa    540
aactacggtc gtgccgtcta cgaatgtctg cgcggcggtc tggacttcac caaagacgac    600
gaaaacatca actcgcagcc gttccaacgc tggcgcgatc gcttcctgtt tgtggctgat    660
gcaatccaca aatcgcaagc agaaaccggt gaaatcaaag gtcactacct gaacgtgacc    720
gcgccgacct gcgaagaaat gatgaaacgg gctgagttcg ctaaagaact cggcatgccg    780
atcatcatgc atgacttctt gacggctggt ttcaccgcca acaccacctt ggcaaaatgg    840
tgccgcgaca acggcgtcct gctgcacatc caccgtgcaa tgcacgcggt gatcgaccgt    900
cagcgtaacc acgggattca cttccgtgtc ttggccaagt gtttgcgtct gtccggtggt    960
gaccacctcc actccggcac cgtcgtcggc aaactggaag gcgacaaagc ttcgaccttg   1020
ggctttgttg acttgatgcg cgaagaccac atcgaagctg accgcagccg tggggtcttc   1080
ttcacccaag attgggcgtc gatgccgggc gtgctgccgg ttgcttccgg tggtatccac   1140
gtgtggcaca tgcccgcact ggtggaaatc ttcggtgatg actccgttct ccagttcggt   1200
ggcggcacct tgggtcaccc ctggggtaat gctcctggtg caaccgcgaa ccgtgttgcc   1260
ttggaagctt gcgtccaagc tcggaacgaa ggtcgcgacc tctaccgtga aggcggcgac   1320
atccttcgtg aagctggcaa gtggtcgcct gaactggctg ctgccctcga cctctggaaa   1380
gagatcaagt tcgaattcga aacgatggac aagctctaag gagcctctga ctatcgctgg   1440
gggagtgagc gttgctgcgt aaagctttct ccccagcctt tcgacttaac ctttcaggat   1500
ttctgaatca tgagcatgaa aactctgccc aaagagcgtc gtttcgagac tttctcgtac   1560
ctgcctcccc tcagcgatcg ccaaatcgct gcacaaatcg agtacatgat cgagcaaggc   1620
ttccaccccc tgatcgagtt caacgagcac tcgaatccgg aagagttcta ctggacgatg   1680
tggaagctcc ccctgtttga ctgcaagagc cctcagcaag tcctcgatga agtgcgtgag   1740
tgccgcagcg aatacggtga ttgctacatc cgtgtcgctg gcttcgacaa catcaagcag   1800
tgccaaaccg tgagcttcat cgttcatcgt cccggccgat actaa                   1845
```

<210> 70
<211> 472
<212> PRT
<213> Artificial

<220>
<223> Com base na sequência traduzida de PCC6301 rbcLS de Synechococcus Sp. do tipo selvagem

<400> 70

```
Met Ala Lys Thr Gln Ser Ala Ala Gly Tyr Lys Ala Gly Val Lys Asp
1               5                  10                  15
Tyr Lys Leu Thr Tyr Tyr Thr Pro Asp Tyr Thr Pro Lys Asp Thr Asp
            20                  25                  30
Leu Leu Ala Ala Phe Arg Phe Ser Pro Gln Pro Gly Val Pro Ala Asp
        35                  40                  45
Glu Ala Gly Ala Ala Ile Ala Ala Glu Ser Ser Thr Gly Thr Trp Thr
    50                  55                  60
Thr Val Trp Thr Asp Leu Leu Thr Asp Met Asp Arg Tyr Lys Gly Lys
65                  70                  75                  80
Cys Tyr His Ile Glu Pro Val Gln Gly Glu Glu Asn Ser Tyr Phe Ala
                85                  90                  95
Phe Ile Ala Tyr Pro Leu Asp Leu Phe Glu Glu Gly Ser Val Thr Asn
            100                 105                 110
Ile Leu Thr Ser Ile Ile Gly Asn Val Phe Gly Phe Lys Ala Ile Arg
        115                 120                 125
Ser Leu Arg Leu Glu Asp Ile Arg Phe Pro Val Ala Leu Val Lys Thr
    130                 135                 140
Phe Gln Gly Pro Pro His Gly Ile Gln Val Glu Arg Asp Leu Leu Asn
145                 150                 155                 160
Lys Tyr Gly Arg Pro Met Leu Gly Cys Thr Ile Lys Pro Lys Leu Gly
                165                 170                 175
Leu Ser Ala Lys Asn Tyr Gly Arg Ala Val Tyr Glu Cys Leu Arg Gly
            180                 185                 190
Gly Leu Asp Phe Thr Lys Asp Asp Glu Asn Ile Asn Ser Gln Pro Phe
        195                 200                 205
Gln Arg Trp Arg Asp Arg Phe Leu Phe Val Ala Asp Ala Ile His Lys
    210                 215                 220
Ser Gln Ala Glu Thr Gly Glu Ile Lys Gly His Tyr Leu Asn Val Thr
225                 230                 235                 240
Ala Pro Thr Cys Glu Glu Met Met Lys Arg Ala Glu Phe Ala Lys Glu
                245                 250                 255
Leu Gly Met Pro Ile Ile Met His Asp Phe Leu Thr Ala Gly Phe Thr
            260                 265                 270
Ala Asn Thr Thr Leu Ala Lys Trp Cys Arg Asp Asn Gly Val Leu Leu
        275                 280                 285
His Ile His Arg Ala Met His Ala Val Ile Asp Arg Gln Arg Asn His
    290                 295                 300
Gly Ile His Phe Arg Val Leu Ala Lys Cys Leu Arg Leu Ser Gly Gly
305                 310                 315                 320
Asp His Leu His Ser Gly Thr Val Val Gly Lys Leu Glu Gly Asp Lys
                325                 330                 335
Ala Ser Thr Leu Gly Phe Val Asp Leu Met Arg Glu Asp His Ile Glu
            340                 345                 350
Ala Asp Arg Ser Arg Gly Val Phe Phe Thr Gln Asp Trp Ala Ser Met
        355                 360                 365
Pro Gly Val Leu Pro Val Ala Ser Gly Gly Ile His Val Trp His Met
    370                 375                 380
Pro Ala Leu Val Glu Ile Phe Gly Asp Asp Ser Val Leu Gln Phe Gly
385                 390                 395                 400
Gly Gly Thr Leu Gly His Pro Trp Gly Asn Ala Pro Gly Ala Thr Ala
                405                 410                 415
Asn Arg Val Ala Leu Glu Ala Cys Val Gln Ala Arg Asn Glu Gly Arg
            420                 425                 430
Asp Leu Tyr Arg Glu Gly Gly Asp Ile Leu Arg Glu Ala Gly Lys Trp
        435                 440                 445
Ser Pro Glu Leu Ala Ala Ala Leu Asp Leu Trp Lys Glu Ile Lys Phe
    450                 455                 460
```

```
Glu Phe Glu Thr Met Asp Lys Leu
465                 470

<210>  71
<211>  1434
<212>  DNA
<213>  Nicotiana tabacum

<400>  71
atgtcaccac aaacagagac taaagcaagt gttggattca aagctggtgt taaagagtac     60
aaattgactt attatactcc tgagtaccaa accaaggata ctgatatatt ggcagcattc    120
cgagtaactc ctcaacctgg agttccacct gaagaagcag gggccgcggt agctgccgaa    180
tcttctactg gtacatggac aactgtatgg accgatggac ttaccagcct tgatcgttac    240
aaagggcgat gctaccgcat cgagcgtgtt gttggagaaa aagatcaata tattgcttat    300
gtagcttacc ctttagacct ttttgaagaa ggttctgtta ccaacatgtt tacttccatt    360
gtaggtaacg tatttgggtt caaagccctg cgcgctctac gtctggaaga tctgcgaatc    420
cctcctgctt atgttaaaac tttccaaggt ccgcctcatg ggatccaagt tgaaagagat    480
aaattgaaca agtatggtcg tcccctgttg ggatgtacta ttaaacctaa attggggtta    540
tctgctaaaa actacggtag agccgtttat gaatgtcttc gcggtggact tgattttact    600
aaagatgatg agaacgtgaa ctcacaacca tttatgcgtt ggagagatcg tttcttattt    660
tgtgccgaag cactttataa agcacaggct gaaacaggtg aaatcaaagg gcattacttg    720
aatgctactg caggtacatg cgaagaaatg atcaaaagag ctgtatttgc tagagaattg    780
ggcgttccga tcgtaatgca tgactactta acgggggat tcaccgcaaa tactagcttg    840
gctcattatt gccgagataa tggtctactt cttcacatcc accgtgcaat gcatgcggtt    900
attgatagac agaagaatca tggtatccac ttccgggtat tagcaaaagc gttacgtatg    960
tctggtggag atcatattca ctctggtacc gtagtaggta aacttgaagg tgaaagagac   1020
ataactttgg gctttgttga tttactgcgt gatgattttg ttgaacaaga tcgaagtcgc   1080
ggtatttatt tcactcaaga ttgggtctct ttaccaggtg ttctacccgt ggcttcagga   1140
ggtattcacg tttggcatat gcctgctctg accgagatct ttggggatga ttccgtacta   1200
cagttcggtg gaggaacttt aggacatcct tggggtaatg cgccaggtgc cgtagctaat   1260
cgagtagctc tagaagcatg tgtaaaagct cgtaatgaag gacgtgatct tgctcaggaa   1320
ggtaatgaaa ttattcgcga ggcttgcaaa tggagcccgg aactagctgc tgcttgtgaa   1380
gtatggaaag agatcgtatt taattttgca gcagtggacg ttttggataa gtaa          1434

<210>  72
<211>  477
<212>  PRT
<213>  Nicotiana tabacum

<400>  72

Met Ser Pro Gln Thr Glu Thr Lys Ala Ser Val Gly Phe Lys Ala Gly
1               5                  10                  15
Val Lys Glu Tyr Lys Leu Thr Tyr Tyr Thr Pro Glu Tyr Gln Thr Lys
            20                  25                  30
Asp Thr Asp Ile Leu Ala Ala Phe Arg Val Thr Pro Gln Pro Gly Val
        35                  40                  45
Pro Pro Glu Glu Ala Gly Ala Ala Val Ala Ala Glu Ser Ser Thr Gly
    50                  55                  60
Thr Trp Thr Thr Val Trp Thr Asp Gly Leu Thr Ser Leu Asp Arg Tyr
65                  70                  75                  80
Lys Gly Arg Cys Tyr Arg Ile Glu Arg Val Val Gly Glu Lys Asp Gln
                85                  90                  95
Tyr Ile Ala Tyr Val Ala Tyr Pro Leu Asp Leu Phe Glu Glu Gly Ser
            100                 105                 110
Val Thr Asn Met Phe Thr Ser Ile Val Gly Asn Val Phe Gly Phe Lys
        115                 120                 125
Ala Leu Arg Ala Leu Arg Leu Glu Asp Leu Arg Ile Pro Pro Ala Tyr
    130                 135                 140
Val Lys Thr Phe Gln Gly Pro Pro His Gly Ile Gln Val Glu Arg Asp
145                 150                 155                 160
```

```
Lys Leu Asn Lys Tyr Gly Arg Pro Leu Leu Gly Cys Thr Ile Lys Pro
                165                 170                 175
Lys Leu Gly Leu Ser Ala Lys Asn Tyr Gly Arg Ala Val Tyr Glu Cys
            180                 185                 190
Leu Arg Gly Gly Leu Asp Phe Thr Lys Asp Asp Glu Asn Val Asn Ser
        195                 200                 205
Gln Pro Phe Met Arg Trp Arg Asp Arg Phe Leu Phe Cys Ala Glu Ala
    210                 215                 220
Leu Tyr Lys Ala Gln Ala Glu Thr Gly Glu Ile Lys Gly His Tyr Leu
225                 230                 235                 240
Asn Ala Thr Ala Gly Thr Cys Glu Glu Met Ile Lys Arg Ala Val Phe
                245                 250                 255
Ala Arg Glu Leu Gly Val Pro Ile Val Met His Asp Tyr Leu Thr Gly
            260                 265                 270
Gly Phe Thr Ala Asn Thr Ser Leu Ala His Tyr Cys Arg Asp Asn Gly
        275                 280                 285
Leu Leu Leu His Ile His Arg Ala Met His Ala Val Ile Asp Arg Gln
    290                 295                 300
Lys Asn His Gly Ile His Phe Arg Val Leu Ala Lys Ala Leu Arg Met
305                 310                 315                 320
Ser Gly Gly Asp His Ile His Ser Gly Thr Val Val Gly Lys Leu Glu
                325                 330                 335
Gly Glu Arg Asp Ile Thr Leu Gly Phe Val Asp Leu Leu Arg Asp Asp
            340                 345                 350
Phe Val Glu Gln Asp Arg Ser Arg Gly Ile Tyr Phe Thr Gln Asp Trp
        355                 360                 365
Val Ser Leu Pro Gly Val Leu Pro Val Ala Ser Gly Gly Ile His Val
    370                 375                 380
Trp His Met Pro Ala Leu Thr Glu Ile Phe Gly Asp Asp Ser Val Leu
385                 390                 395                 400
Gln Phe Gly Gly Gly Thr Leu Gly His Pro Trp Gly Asn Ala Pro Gly
                405                 410                 415
Ala Val Ala Asn Arg Val Ala Leu Glu Ala Cys Val Lys Ala Arg Asn
            420                 425                 430
Glu Gly Arg Asp Leu Ala Gln Glu Gly Asn Glu Ile Ile Arg Glu Ala
        435                 440                 445
Cys Lys Trp Ser Pro Glu Leu Ala Ala Ala Cys Glu Val Trp Lys Glu
    450                 455                 460
Ile Val Phe Asn Phe Ala Ala Val Asp Val Leu Asp Lys
465                 470                 475
```

<210> 73
<211> 1434
<212> DNA
<213> Artificial

<220>
<223> Com base na sequência de codificações de rbcL de Nicotina tabacum do tipo selvagem

<400> 73

```
atgtcaccac aaacagagac taaagcaagt gttggattca aagctggtgt taaagagtac     60
aaattgactt attatactcc tgagtaccaa accaaggata ctgatatatt ggcagcattc    120
cgagtaactc ctcaacctgg agttccacct gaagaagcag gggccgcggt agctgccgaa    180
tcttctactg gtacatggac aactgtatgg accgatggac ttaccagcct tgatcgttac    240
agagggcgat gctaccgcat cgagcgtgtt gttggagaaa aagatcaata tattgcttat    300
gtagcttacc ctttagacct ttttgaagaa ggttctgtta ccaacatgtt tacttccatt    360
gtaggtaacg tatttgggtt caaagccctg cgcgctctac gtctggaaga tctgcgaatc    420
cctcctgctt atgttaaaac tttccaaggt ccgcctcatg ggatccaagt tgaaagagat    480
aaattgaaca agtatggtcg tcccctgttg ggatgtacta ttaaacctaa attggggtta    540
tctgctaaaa actacggtag agccgtttat gaatgtcttc gcggtggact tgattttact    600
aaagatgatg agaacgtgaa ctcacaacca tttatgcgtt ggagagatcg tttcttattt    660
```

```
tgtgccgaag cactttataa agcacaggct gaaacaggtg aaatcaaagg gcattacttg    720
aatgctactg caggtacatg cgaagaaatg atcaaaagag ctgtatttgc tagagaattg    780
ggcgttccga tcgtaatgca tgactactta acgggggggat tcaccgcaaa tactagcttg    840
gctcattatt gccgagataa tggtctactt cttcacatcc accgtgcaat gcatgcggtt    900
attgatagac agaagaatca tggtatccac ttccgggtat tagcaaaagc gttacgtatg    960
tctggtggag atcatattca ctctggtacc gtagtaggta aacttgaagg tgaaagagac   1020
ataactttgg gctttgttga tttactgcgt gatgattttg ttgaacaaga tcgaagtcgc   1080
ggtatttatt tcactcaaga ttgggtctct ttaccaggtg ttctacccgt ggcttcagga   1140
ggtattcacg tttggcatat gcctgctctg accgagatct ttggggatga ttccgtacta   1200
cagttcggtg gaggaacttt aggacatcct tggggtaatg cgccaggtgc cgtagctaat   1260
cgagtagctc tagaagcatg tgtaaaagct cgtaatgaag gacgtgatct tgctcaggaa   1320
ggtaatgaaa ttattcgcga ggcttgcaaa tggagcccgg aactagctgc tgcttgtgaa   1380
gtatggaaag agatcgtatt taattttgca gcagtggacg ttttggataa gtaa          1434
```

<210> 74
<211> 477
<212> PRT
<213> Artificial

<220>
<223> Com base na sequência de subunidade L traduzida do tipo selvagem de icotina tabacum do tipo selvagem

<400> 74

```
Met Ser Pro Gln Thr Glu Thr Lys Ala Ser Val Gly Phe Lys Ala Gly
1               5                  10                  15
Val Lys Glu Tyr Lys Leu Gly Tyr Tyr Thr Pro Glu Tyr Gln Thr Lys
            20                  25                  30
Asp Thr Asp Ile Leu Ala Ala Phe Arg Val Thr Pro Gln Pro Gly Val
        35                  40                  45
Pro Pro Glu Glu Ala Gly Ala Ala Val Ala Ala Glu Ser Ser Thr Gly
    50                  55                  60
Thr Trp Thr Thr Val Trp Thr Asp Gly Leu Thr Ser Leu Asp Arg Tyr
65                  70                  75                  80
Arg Gly Arg Cys Tyr Arg Ile Glu Arg Val Val Gly Glu Lys Asp Gln
                85                  90                  95
Tyr Ile Ala Tyr Val Ala Tyr Pro Leu Asp Leu Phe Glu Glu Gly Ser
            100                 105                 110
Val Thr Asn Met Phe Thr Ser Ile Val Gly Asn Val Phe Gly Phe Lys
        115                 120                 125
Ala Leu Arg Ala Leu Arg Leu Glu Asp Leu Arg Ile Pro Pro Ala Tyr
    130                 135                 140
Val Lys Thr Phe Gln Gly Pro Pro His Gly Ile Gln Val Glu Arg Asp
145                 150                 155                 160
Lys Leu Asn Lys Tyr Gly Arg Pro Leu Leu Gly Cys Thr Ile Lys Pro
                165                 170                 175
Lys Leu Gly Leu Ser Ala Lys Asn Tyr Gly Arg Ala Val Tyr Glu Cys
            180                 185                 190
Leu Arg Gly Gly Leu Asp Phe Thr Lys Asp Asp Glu Asn Val Asn Ser
        195                 200                 205
Gln Pro Phe Met Arg Trp Arg Asp Arg Phe Leu Phe Cys Ala Glu Ala
    210                 215                 220
Leu Tyr Lys Ala Gln Ala Glu Thr Gly Glu Ile Lys Gly His Tyr Leu
225                 230                 235                 240
Asn Ala Thr Ala Gly Thr Cys Glu Glu Met Ile Lys Arg Ala Val Phe
                245                 250                 255
Ala Arg Glu Leu Gly Val Pro Ile Val Met His Asp Tyr Leu Thr Gly
            260                 265                 270
Gly Phe Thr Ala Asn Thr Ser Leu Ala His Tyr Cys Arg Asp Asn Gly
        275                 280                 285
```

```
Leu Leu Leu His Ile His Arg Ala Met His Ala Val Ile Asp Arg Gln
    290                 295                 300
Lys Asn His Gly Ile His Phe Arg Val Leu Ala Lys Ala Leu Arg Met
305                 310                 315                 320
Ser Gly Gly Asp His Ile His Ser Gly Thr Val Val Gly Lys Leu Glu
                325                 330                 335
Gly Glu Arg Asp Ile Thr Leu Gly Phe Val Asp Leu Leu Arg Asp Asp
            340                 345                 350
Phe Val Glu Gln Asp Arg Ser Arg Gly Ile Tyr Phe Thr Gln Asp Trp
        355                 360                 365
Val Ser Leu Pro Gly Val Leu Pro Val Ala Ser Gly Gly Ile His Val
    370                 375                 380
Trp His Met Pro Ala Leu Thr Glu Ile Phe Gly Asp Asp Ser Val Leu
385                 390                 395                 400
Gln Phe Gly Gly Gly Thr Leu Gly His Pro Trp Gly Asn Ala Pro Gly
                405                 410                 415
Ala Val Ala Asn Arg Val Ala Leu Glu Ala Cys Val Lys Ala Arg Asn
            420                 425                 430
Glu Gly Arg Asp Leu Ala Gln Glu Gly Asn Glu Ile Ile Arg Glu Ala
        435                 440                 445
Cys Lys Trp Ser Pro Glu Leu Ala Ala Ala Cys Glu Val Trp Lys Glu
    450                 455                 460
Ile Val Phe Asn Phe Ala Ala Val Asp Val Leu Asp Lys
465                 470                 475
```

<210> 75
<211> 1434
<212> DNA
<213> Artificial

<220>
<223> Com base na sequência de codificações de rbcL de Nicotina tabacum do tipo selvagem

<400> 75

```
atgtcaccac aaacagagac taaagcaagt gttggattca aagctggtgt taaagagtac     60
aaattgactt attatactcc tgagtaccaa accaaggata ctgatatatt ggcagcattc    120
cgagtaactc ctcaacctgg agttccacct gaagaagcag gggccgcggt agctgccgaa    180
tcttctactg gtacatggac aactgtatgg accgatggac ttaccagcct tgatcgttac    240
agagggcgat gctaccgcat cgagcgtgtt gttggagaaa aagatcaata tattgcttat    300
gtagcttacc ctttagacct ttttgaagaa ggttctgtta ccaacatgtt tacttccatt    360
gtaggtaacg tatttgggtt caaagccctg cgcgctctac gtctggaaga tctgcgaatc    420
cctcctgctt atgttaaaac tttccaaggt ccgcctcatg ggatccaagt tgaaagagat    480
aaattgaaca agtatggtcg tcccctgttg ggatgtacta ttaaacctaa attggggtta    540
tctgctaaaa actacggtag agccgtttat gaatgtcttc gcggtggact tgattttact    600
aaagatgatg agaacgtgaa ctcacaacca tttatgcgtt ggagagatcg tttcttattt    660
tgtgccgaag cactttataa agcacaggct gaaacaggtg aaatcaaagg gcattacttg    720
aatgctactg caggtacatg cgaagaaatg atcaaaagag ctgtatttgc tagagaattg    780
ggcgttccga tcgtaatgca tgactactta acgggggat tcaccgcaaa tactagcttg    840
gctcattatt gccgagataa tggtctactt cttcacatcc accgtgcaat gcatgcggtt    900
attgatagac agaagaatca tggtatccac ttccgggtat tagcaaaagc gttacgtatg    960
tctggtggag atcatattca ctctggtacc gtagtaggta aacttgaagg tgaaagagac   1020
ataactttgg gctttgttga tttactgcgt gatgattttg ttgaacaaga tcgaagtcgc   1080
ggtatttatt tcactcaaga ttgggtctct ttaccaggtg ttctacccgt ggcttcagga   1140
ggtattcacg tttggcatat gcctgctctg accgagatct ttggggatga ttccgtacta   1200
cagttcggtg gaggaacttt aggacatcct tggggtaatg cgccaggtgc cgtagctaat   1260
cgagtagctc tagaagcatg tgtaaaagct cgtaatgaag gacgtgatct tgctcaggaa   1320
ggtaatgaaa ttattcgcga ggcttgcaaa tggagcccgg aactagctgc tgcttgtgaa   1380
gtatggaaag agatcgtatt taattttgca gcagtggacg ttttggataa gtaa          1434
```

<210> 76
<211> 477

<212> PRT
<213> Artificial

<220>
<223> Com base na subunidade L traduzida de Nicotina tabacum do tipo selvagem

<400> 76

```
Met Ser Pro Gln Thr Glu Thr Lys Ala Ser Val Gly Phe Lys Ala Gly
1               5                  10                  15
Val Lys Glu Tyr Lys Leu Thr Tyr Tyr Thr Pro Glu Tyr Gln Thr Lys
            20                  25                  30
Asp Thr Asp Ile Leu Ala Ala Phe Arg Val Thr Pro Gln Pro Gly Val
        35                  40                  45
Pro Pro Glu Glu Ala Gly Ala Ala Val Ala Ala Glu Ser Ser Thr Gly
    50                  55                  60
Thr Trp Thr Thr Val Trp Thr Asp Gly Leu Thr Ser Leu Asp Arg Tyr
65                  70                  75                  80
Arg Gly Arg Cys Tyr Arg Ile Glu Arg Val Val Gly Glu Lys Asp Gln
                85                  90                  95
Tyr Ile Ala Tyr Val Ala Tyr Pro Leu Asp Leu Phe Glu Glu Gly Ser
            100                 105                 110
Val Thr Asn Met Phe Thr Ser Ile Val Gly Asn Val Phe Gly Phe Lys
        115                 120                 125
Ala Leu Arg Ala Leu Arg Leu Glu Asp Leu Arg Ile Pro Pro Ala Tyr
    130                 135                 140
Val Lys Thr Phe Gln Gly Pro Pro His Gly Ile Gln Val Glu Arg Asp
145                 150                 155                 160
Lys Leu Asn Lys Tyr Gly Arg Pro Leu Leu Gly Cys Thr Ile Lys Pro
                165                 170                 175
Lys Leu Gly Leu Ser Ala Lys Asn Tyr Gly Arg Ala Val Tyr Glu Cys
            180                 185                 190
Leu Arg Gly Gly Leu Asp Phe Thr Lys Asp Asp Glu Asn Val Asn Ser
        195                 200                 205
Gln Pro Phe Met Arg Trp Arg Asp Arg Phe Leu Phe Cys Ala Glu Ala
    210                 215                 220
Leu Tyr Lys Ala Gln Ala Glu Thr Gly Glu Ile Lys Gly His Tyr Leu
225                 230                 235                 240
Asn Ala Thr Ala Gly Thr Cys Glu Glu Met Ile Lys Arg Ala Val Phe
                245                 250                 255
Ala Arg Glu Leu Gly Val Pro Ile Val Met His Asp Tyr Leu Thr Gly
            260                 265                 270
Gly Phe Thr Ala Asn Thr Ser Leu Ala His Tyr Cys Arg Asp Asn Gly
        275                 280                 285
Leu Leu Leu His Ile His Arg Ala Met His Ala Val Ile Asp Arg Gln
    290                 295                 300
Lys Asn His Gly Ile His Phe Arg Val Leu Ala Lys Ala Leu Arg Met
305                 310                 315                 320
Ser Gly Gly Asp His Ile His Ser Gly Thr Val Val Gly Lys Leu Glu
                325                 330                 335
Gly Glu Arg Asp Ile Thr Leu Gly Phe Val Asp Leu Leu Arg Asp Asp
            340                 345                 350
Phe Val Glu Gln Asp Arg Ser Arg Gly Ile Tyr Phe Thr Gln Asp Trp
        355                 360                 365
Val Ser Leu Pro Gly Val Leu Pro Val Ala Ser Gly Gly Ile His Val
    370                 375                 380
Trp His Met Pro Ala Leu Thr Glu Ile Phe Gly Asp Asp Ser Val Leu
385                 390                 395                 400
Gln Phe Gly Gly Gly Thr Leu Gly His Pro Trp Gly Asn Ala Pro Gly
                405                 410                 415
Ala Val Ala Asn Arg Val Ala Leu Glu Ala Cys Val Lys Ala Arg Asn
```

```
            420                 425                 430
Glu Gly Arg Asp Leu Ala Gln Glu Gly Asn Glu Ile Ile Arg Glu Ala
        435                 440                 445
Cys Lys Trp Ser Pro Glu Leu Ala Ala Ala Cys Glu Val Trp Lys Glu
    450                 455                 460
Ile Val Phe Asn Phe Ala Ala Val Asp Val Leu Asp Lys
465                 470                 475
```

<210> 77
<211> 1434
<212> DNA
<213> Artificial

<220>
<223> Com base na sequência de codificações de rbcL de Nicotina tabacum do tipo selvagem

<400> 77

```
atgtcaccac aaacagagac taaagcaagt gttggattca aagctggtgt taaagagtac     60
aaattgggtt attatactcc tgagtaccaa accaaggata ctgatatatt ggcagcattc    120
cgagtaactc ctcaacctgg agttccacct gaagaagcag gggccgcggt agctgccgaa    180
tcttctactg gtacatggac aactgtatgg accgatggac ttaccagcct tgatcgttac    240
aaagggcgat gctaccgcat cgagcgtgtt gttggagaaa aagatcaata tattgcttat    300
gtagcttacc ctttagacct ttttgaagaa ggttctgtta ccaacatgtt tacttccatt    360
gtaggtaacg tatttgggtt caaagccctg cgcgctctac gtctggaaga tctgcgaatc    420
cctcctgctt atgttaaaac tttccaaggt ccgcctcatg ggatccaagt tgaaagagat    480
aaattgaaca agtatggtcg tcccctgttg ggatgtacta ttaaacctaa attggggtta    540
tctgctaaaa actacggtag agccgtttat gaatgtcttc gcggtggact tgattttact    600
aaagatgatg agaacgtgaa ctcacaacca tttatgcgtt ggagagatcg tttcttattt    660
tgtgccgaag cactttataa agcacaggct gaaacaggtg aaatcaaagg gcattacttg    720
aatgctactg caggtacatg cgaagaaatg atcaaaagag ctgtatttgc tagagaattg    780
ggcgttccga tcgtaatgca tgactactta acgggggat tcaccgcaaa tactagcttg    840
gctcattatt gccgagataa tggtctactt cttcacatcc accgtgcaat gcatgcggtt    900
attgatagac agaagaatca tggtatccac ttccgggtat tagcaaaagc gttacgtatg    960
tctggtggag atcatattca ctctggtacc gtagtaggta aacttgaagg tgaaagagac   1020
ataactttgg gctttgttga tttactgcgt gatgattttg ttgaacaaga tcgaagtcgc   1080
ggtatttatt tcactcaaga ttgggtctct ttaccaggtg ttctacccgt ggcttcagga   1140
ggtattcacg tttggcatat gcctgctctg accgagatct ttggggatga ttccgtacta   1200
cagttcggtg gaggaacttt aggacatcct tggggtaatg cgccaggtgc cgtagctaat   1260
cgagtagctc tagaagcatg tgtaaaagct cgtaatgaag gacgtgatct tgctcaggaa   1320
ggtaatgaaa ttattcgcga ggcttgcaaa tggagcccgg aactagctgc tgcttgtgaa   1380
gtatggaaag agatcgtatt taattttgca gcagtggacg ttttggataa gtaa          1434
```

<210> 78
<211> 477
<212> PRT
<213> Artificial

<220>
<223> Com base na subunidade L traduzida de Nicotina tabacum do tipo selvagem

<400> 78

```
Met Ser Pro Gln Thr Glu Thr Lys Ala Ser Val Gly Phe Lys Ala Gly
1               5                  10                  15
Val Lys Glu Tyr Lys Leu Gly Tyr Tyr Thr Pro Glu Tyr Gln Thr Lys
            20                  25                  30
Asp Thr Asp Ile Leu Ala Ala Phe Arg Val Thr Pro Gln Pro Gly Val
        35                  40                  45
Pro Pro Glu Glu Ala Gly Ala Ala Val Ala Ala Glu Ser Ser Thr Gly
    50                  55                  60
```

```
Thr Trp Thr Thr Val Trp Thr Asp Gly Leu Thr Ser Leu Asp Arg Tyr
65                  70                  75                  80
Lys Gly Arg Cys Tyr Arg Ile Glu Arg Val Val Gly Glu Lys Asp Gln
                85                  90                  95
Tyr Ile Ala Tyr Val Ala Tyr Pro Leu Asp Leu Phe Glu Glu Gly Ser
            100                 105                 110
Val Thr Asn Met Phe Thr Ser Ile Val Gly Asn Val Phe Gly Phe Lys
        115                 120                 125
Ala Leu Arg Ala Leu Arg Leu Glu Asp Leu Arg Ile Pro Pro Ala Tyr
    130                 135                 140
Val Lys Thr Phe Gln Gly Pro Pro His Gly Ile Gln Val Glu Arg Asp
145                 150                 155                 160
Lys Leu Asn Lys Tyr Gly Arg Pro Leu Leu Gly Cys Thr Ile Lys Pro
                165                 170                 175
Lys Leu Gly Leu Ser Ala Lys Asn Tyr Gly Arg Ala Val Tyr Glu Cys
            180                 185                 190
Leu Arg Gly Gly Leu Asp Phe Thr Lys Asp Asp Glu Asn Val Asn Ser
        195                 200                 205
Gln Pro Phe Met Arg Trp Arg Asp Arg Phe Leu Phe Cys Ala Glu Ala
    210                 215                 220
Leu Tyr Lys Ala Gln Ala Glu Thr Gly Glu Ile Lys Gly His Tyr Leu
225                 230                 235                 240
Asn Ala Thr Ala Gly Thr Cys Glu Glu Met Ile Lys Arg Ala Val Phe
                245                 250                 255
Ala Arg Glu Leu Gly Val Pro Ile Val Met His Asp Tyr Leu Thr Gly
            260                 265                 270
Gly Phe Thr Ala Asn Thr Ser Leu Ala His Tyr Cys Arg Asp Asn Gly
        275                 280                 285
Leu Leu Leu His Ile His Arg Ala Met His Ala Val Ile Asp Arg Gln
    290                 295                 300
Lys Asn His Gly Ile His Phe Arg Val Leu Ala Lys Ala Leu Arg Met
305                 310                 315                 320
Ser Gly Gly Asp His Ile His Ser Gly Thr Val Val Gly Lys Leu Glu
                325                 330                 335
Gly Glu Arg Asp Ile Thr Leu Gly Phe Val Asp Leu Leu Arg Asp Asp
            340                 345                 350
Phe Val Glu Gln Asp Arg Ser Arg Gly Ile Tyr Phe Thr Gln Asp Trp
        355                 360                 365
Val Ser Leu Pro Gly Val Leu Pro Val Ala Ser Gly Gly Ile His Val
    370                 375                 380
Trp His Met Pro Ala Leu Thr Glu Ile Phe Gly Asp Asp Ser Val Leu
385                 390                 395                 400
Gln Phe Gly Gly Gly Thr Leu Gly His Pro Trp Gly Asn Ala Pro Gly
                405                 410                 415
Ala Val Ala Asn Arg Val Ala Leu Glu Ala Cys Val Lys Ala Arg Asn
            420                 425                 430
Glu Gly Arg Asp Leu Ala Gln Glu Gly Asn Glu Ile Ile Arg Glu Ala
        435                 440                 445
Cys Lys Trp Ser Pro Glu Leu Ala Ala Ala Cys Glu Val Trp Lys Glu
    450                 455                 460
Ile Val Phe Asn Phe Ala Ala Val Asp Val Leu Asp Lys
465                 470                 475
```

REIVINDICAÇÕES

1. Método para gerar uma proteína Rubisco com uma propriedade funcional melhorada selecionada dentre o grupo que consiste em maior eficiência cinética da Rubisco, uma especificidade alterada da Rubisco para um ou mais substratos, uma especificidade alterada para um ou mais produtos da Rubisco e uma faixa de temperatura efetiva alterada para a catálise da Rubisco, caracterizado pelo fato de que compreende:

(a) identificar pelo menos um resíduo de aminoácidos alvo em uma primeira proteína Rubisco, em que o referido resíduo de aminoácidos alvo está associado à referida propriedade funcional, em que os resíduos de aminoácidos alvo da proteína Rubisco são resíduos de "primeira concha" que coordenam diretamente com o centro de reação ou resíduos de "segunda concha" que coordenam diretamente com um ou mais resíduos da "primeira concha", em que os resíduos da primeira concha de Rubisco são selecionados dentre o grupo que consiste em GLU60, ASN123, LYS175, LYS177, KCX201, ASP203, GLU204, HIS294, em que a numeração de resíduos é de Rubisco de espinafre e KCX201 indica LYS201 carbamilado;

(b) comparar pelo menos uma segunda proteína Rubisco homóloga da mesma ramificação filogenética ou de uma diferente que a primeira proteína Rubisco com a primeira proteína Rubisco e identificar pelo menos um resíduo de aminoácido variante entre a primeira proteína Rubisco e a segunda proteína Rubisco, em que os resíduos da segunda proteína Rubisco utilizados para comparação com a primeira proteína Rubisco compreendem todos os resíduos diretamente coordenados ao sítio ativo de uma proteína Rubisco, todos os resíduos interagindo com o centro reativo do substrato ou com as espécies de reação intermediária de uma proteína Rubisco e outros resíduos a uma distância aproximada entre 3 e 28 Å de qualquer átomo de subs-

trato ou espécie de reação intermediária do sítio ativo de uma proteína Rubisco ou de um subconjunto desses resíduos;

(c) selecionar pelo menos um resíduo de aminoácido candidato dos resíduos de aminoácido variantes identificados em (b) com base na avaliação da proximidade espacial dos resíduos variantes aos resíduos alvo e estimar e classificar sua capacidade de influenciar a eletrostática e a orientação dos resíduos alvo e modular a propriedade funcional da proteína formada;

(d) produzir pelo menos uma proteína Rubisco mutante candidata *in vitro* em que o referido pelo menos um resíduo de aminoácido candidato da segunda proteína Rubisco substitui um resíduo correspondente na primeira proteína Rubisco; e

(e) triar a referida pelo menos uma proteína Rubisco mutante candidata produzida em (d) para identificar uma proteína Rubisco tendo a referida propriedade funcional melhorada.

2. Método, de acordo com a reivindicação 1, caracterizado pelo fato de que a etapa (a) e a etapa (b) são executadas simultaneamente.

3. Método, de acordo com a reivindicação 1, caracterizado pelo fato de que a etapa (b) é executada antes da etapa (a).

4. Método, de acordo com qualquer uma das reivindicações 1 a 3, caracterizado pelo fato de que a etapa (d) compreende produzir pelo menos uma proteína Rubisco mutante candidata usando o pelo menos um resíduo de aminoácido candidato da segunda proteína para substituir um resíduo correspondente em uma proteína Rubisco homóloga ou em proteínas Rubisco homólogas diferentes ou além da primeira proteína Rubisco.

5. Método, de acordo com qualquer uma das reivindicações 1 a 4, caracterizado pelo fato de que a etapa (d) compreende produzir pelo menos uma proteína Rubisco mutante candidata em que pelo

menos dois resíduos de aminoácido candidatos da segunda proteína Rubisco substituem resíduos correspondentes na primeira proteína Rubisco e/ou em uma proteína homóloga diferente da primeira proteína Rubisco.

6. Método, de acordo com qualquer uma das reivindicações 1 a 5, caracterizado pelo fato de que o pelo menos um resíduo de aminoácidos alvo está contido em um conjunto de proteínas contendo a primeira proteína Rubisco.

7. Método, de acordo com qualquer uma das reivindicações 1 a 6, caracterizado pelo fato de que o pelo menos um resíduo de aminoácidos alvo é pelo menos 2, pelo menos 3, pelo menos 4, pelo menos 5, pelo menos 10, pelo menos 12, pelo menos 15, pelo menos 20, pelo menos 30, ou pelo menos 50 resíduos de aminoácidos alvos.

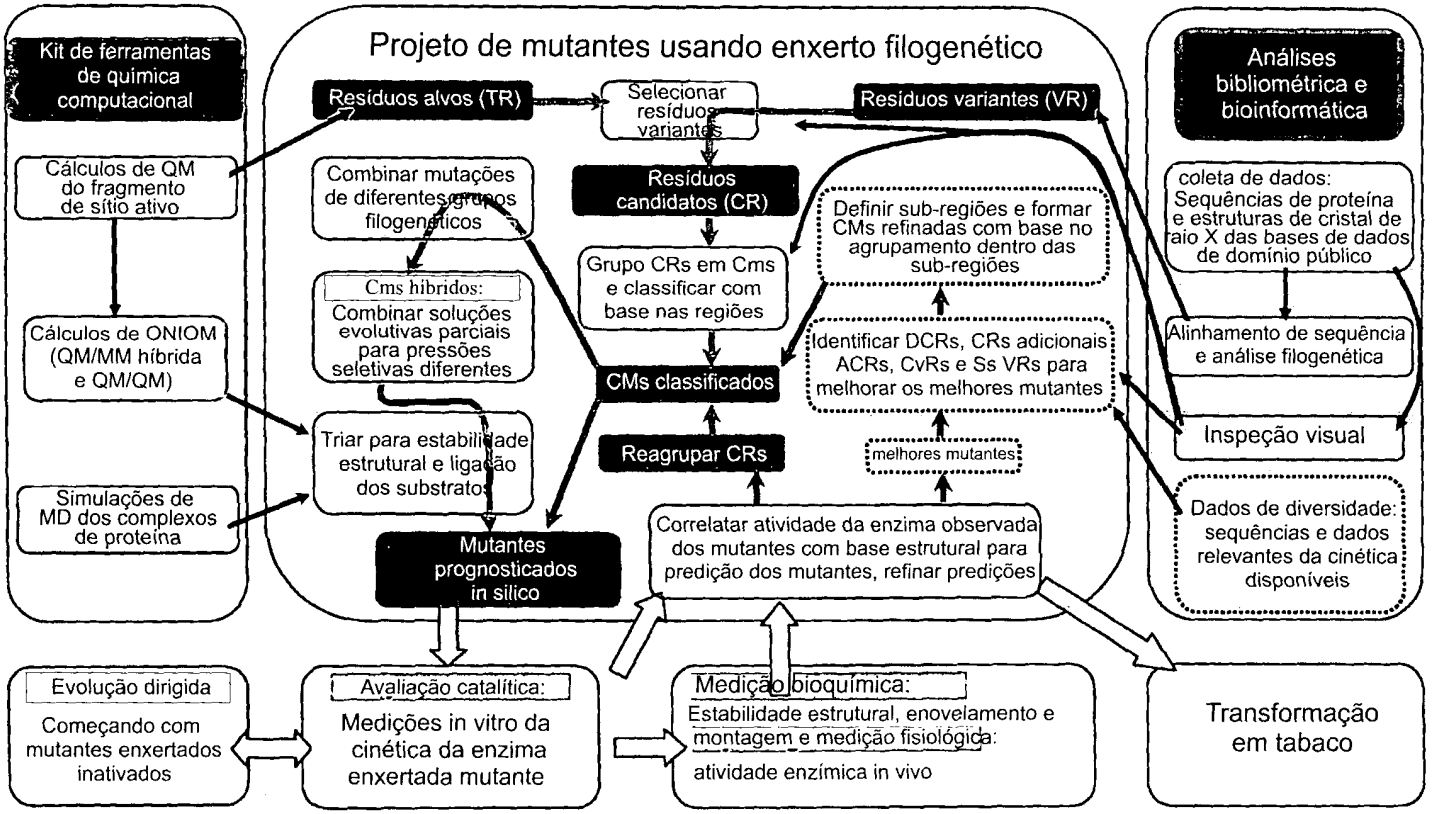
8. Método, de acordo com a reivindicação 1, caracterizado pelo fato de que os resíduos de aminoácidos alvos da proteína de Rubisco encontram-se no domínio N-terminal da subunidade grande da Rubisco.

9. Método, de acordo com qualquer uma das reivindicações 1 a 8, caracterizado pelo fato de que a etapa (c) compreende selecionar pelo menos um resíduo de aminoácido candidato divergente e/ou pelo menos um resíduo de aminoácido candidato alternativo, em vez de pelo menos um resíduo de aminoácido candidato, dos resíduos de aminoácido variantes identificados em (b), em que o referido resíduo de aminoácido candidato divergente ou resíduo candidato divergente (DCR) refere-se a um resíduo de aminoácido que é selecionado dentre uma pluralidade de resíduos variantes e que é suspeito de poder influenciar um ou mais resíduos alvo estereoquimicamente e/ou eletrostaticamente, e assim influenciar a função da proteína Rubisco mediada por um ou mais resíduos alvo; e em que o referido resíduo de aminoácido candidato alternativo ou resíduo candiado alternativo (ACR) refe-

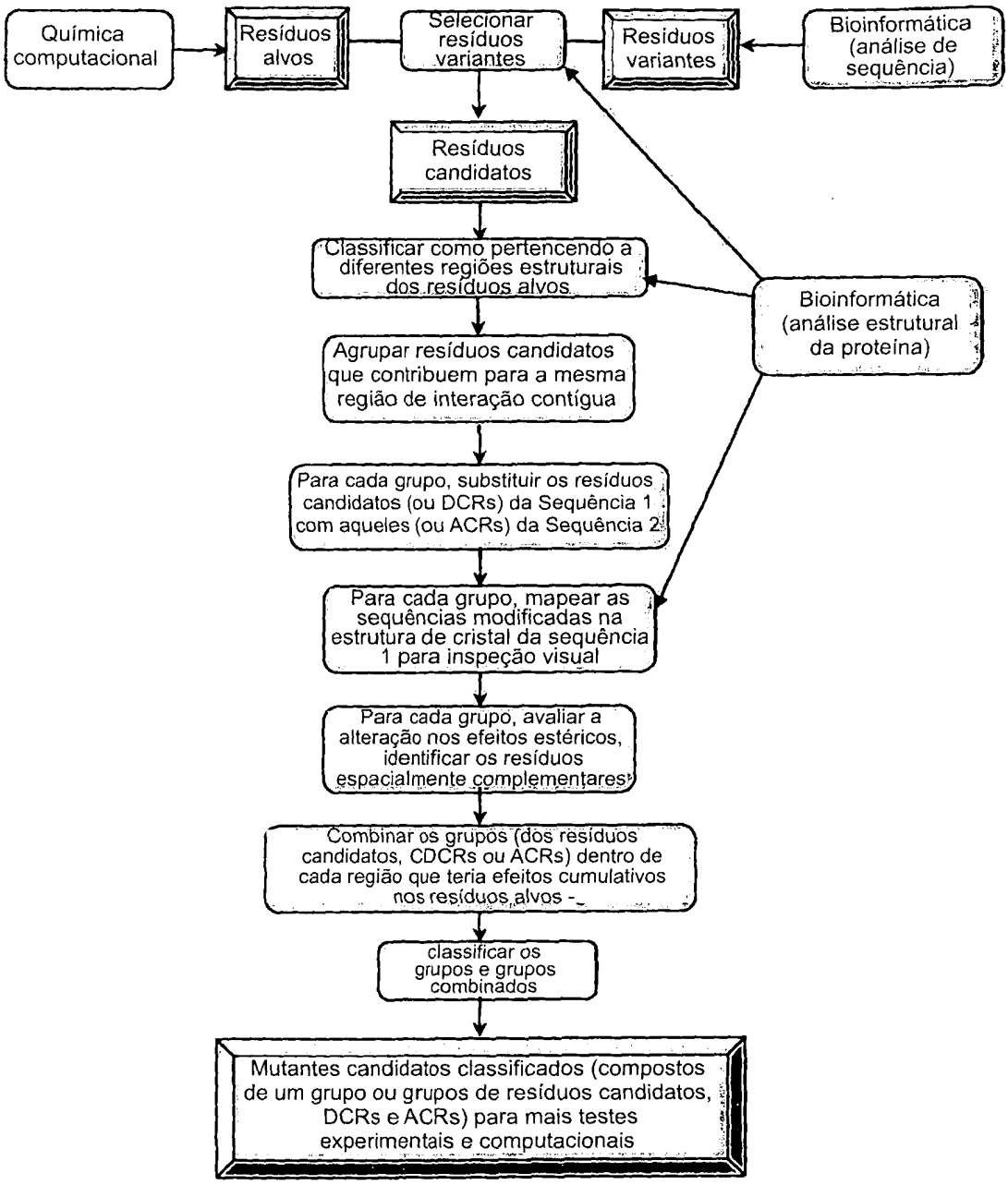
re-se a um resíduo de aminoácido alternativo na posição de um resíduo candidato que é expresso em uma segunda proteína Rubisco, mas que não é o aminoácido expresso na sequência consenso da segunda proteína Rubisco e que é suspeita de poder influenciar um ou mais resíduos alvo estereoquimicamente e/ou eletrostaticamente e, assim, influenciar a função da proteína Rubisco mediada por um ou mais resíduos alvo.

10. Método de acordo com qualquer uma das reivindicações 1 a 9, caracterizado pelo fato de que compreende a etapa adicional de executar evolução direcionada na referida proteína Rubisco tendo a referida propriedade funcional melhorada e produtos de triagem da mesma.

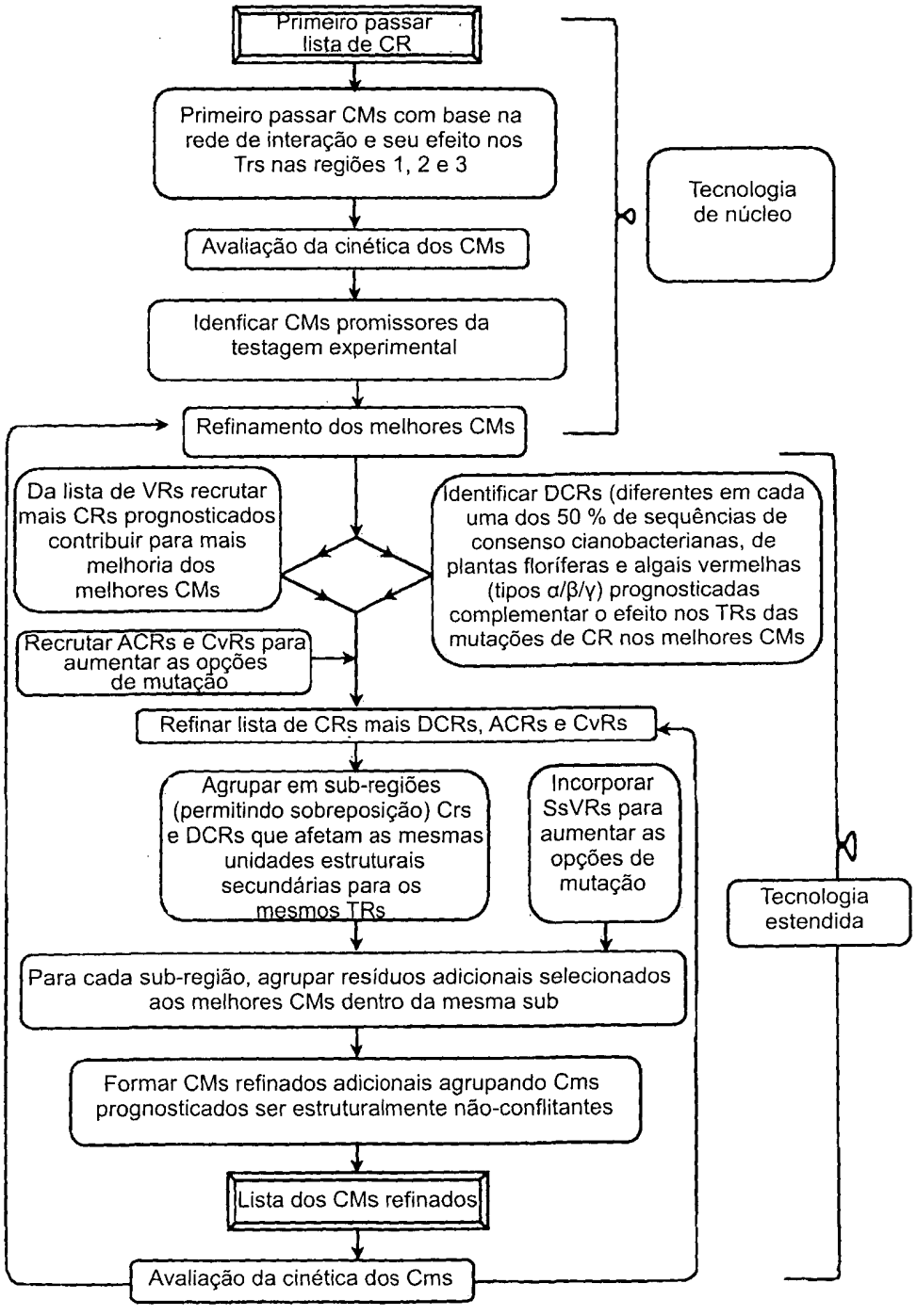
Estratégia de Reengenharia de Rubisco Refinada



Fluxograma 1



Fluxograma 2



Fluxograma 3

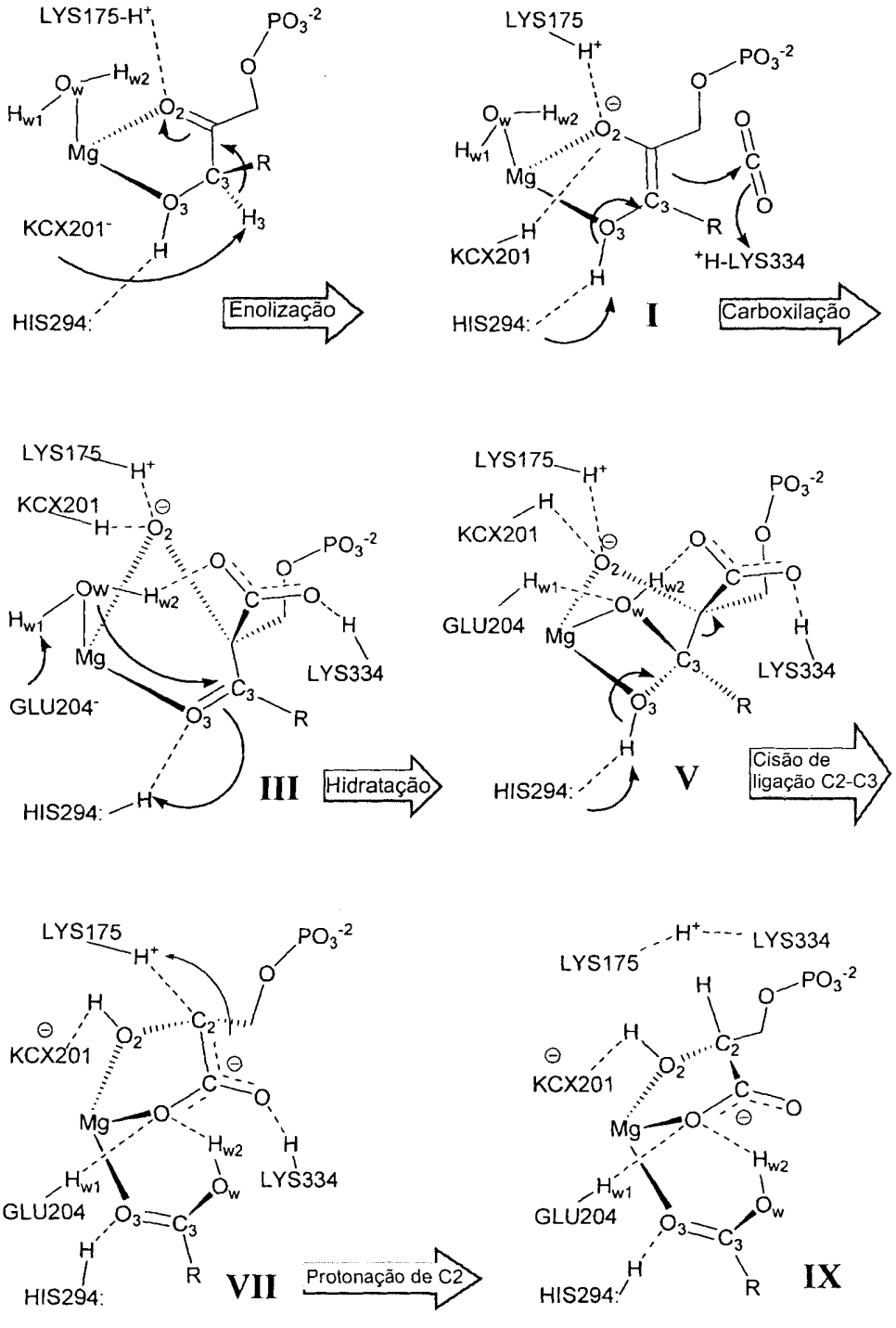

LYS175-H+
PO3-2
Ow
Hw2
Hw1
O2
Mg
C3
R
O3
KCX201-
H3
H
HIS294:
Enolização
LYS175
H+
KCX201
I
Carboxilação
+H-LYS334
III
Hidratação
LYS334
GLU204-
V
Cisão de ligação C2-C3
GLU204
VII
Protonação de C2
IX
R = -CH(OH)-CH2-O-PO3 2-

FIG. 1

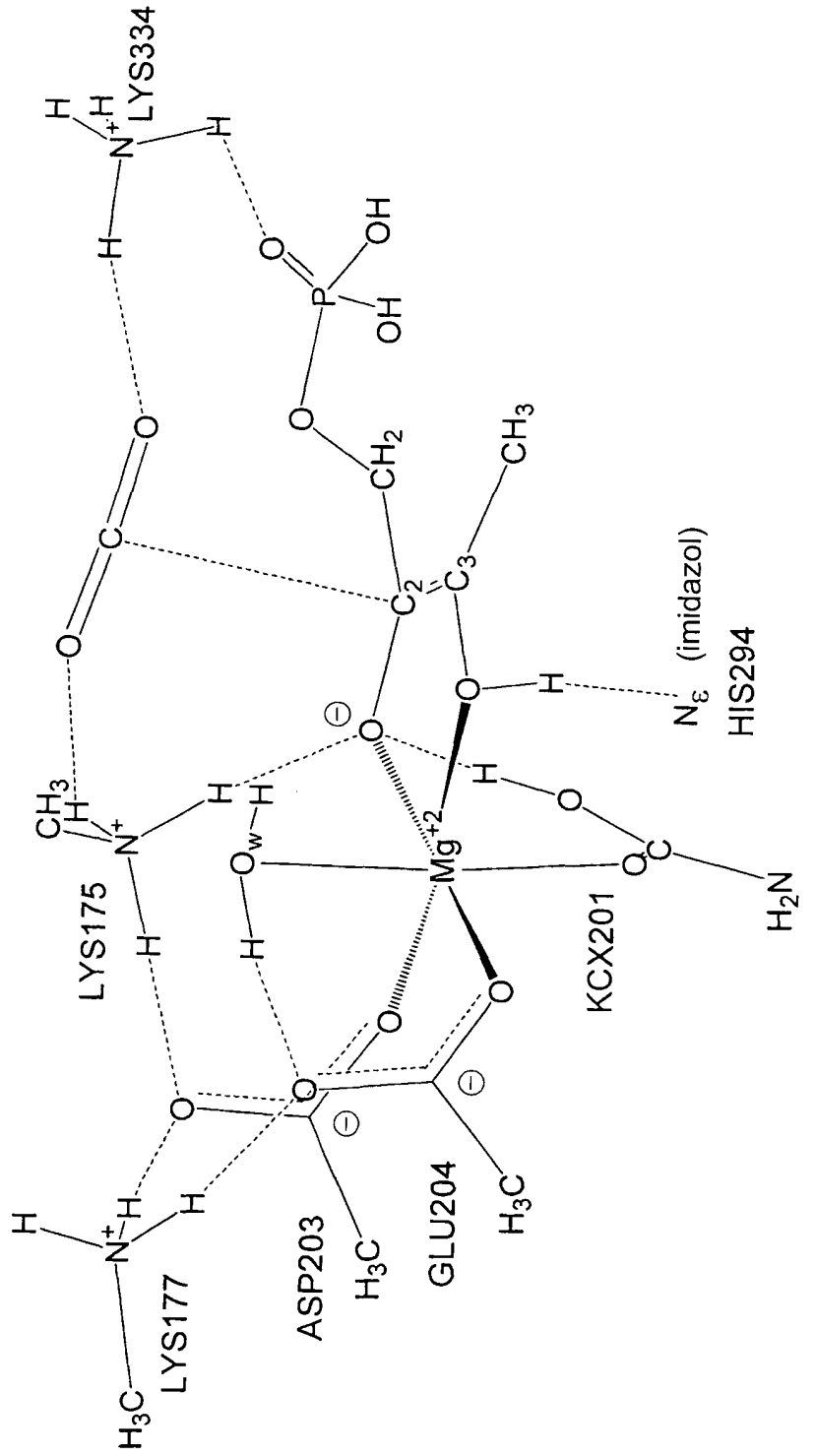

LYS334
LYS175
LYS177
ASP203
GLU204
KCX201
HIS294
(imidazol)

FIG. 2

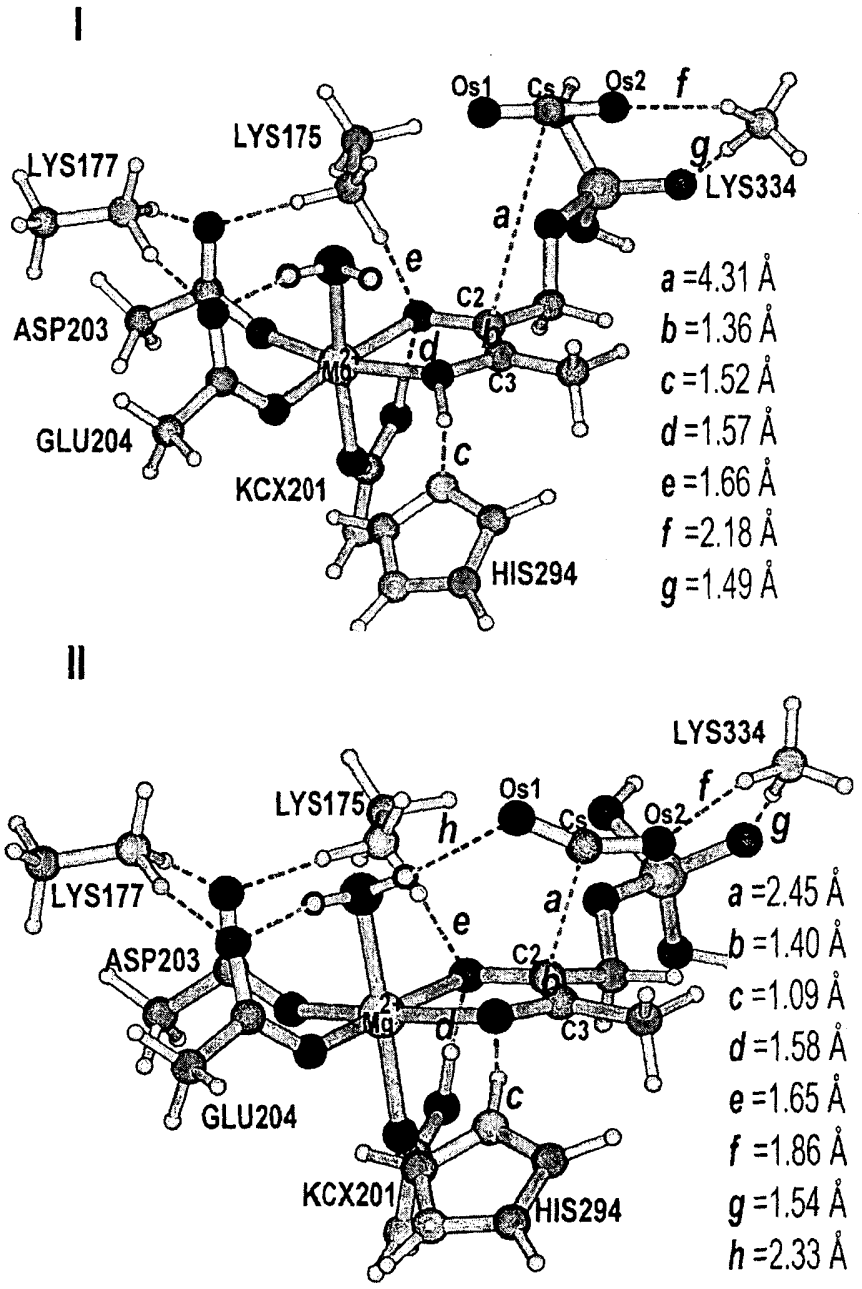

I
Os1
Cs
Os2
f
g
LYS334
LYS175
LYS177
a
e
C2
b
ASP203
d
Mg
C3
GLU204
c
KCX201
HIS294
a =4.31 Å
b =1.36 Å
c =1.52 Å
d =1.57 Å
e =1.66 Å
f =2.18 Å
g =1.49 Å
II
LYS334
LYS175
h
Os1
Cs
Os2
f
g
LYS177
a
e
ASP203
C2
b
Mg
d
C3
GLU204
c
KCX201
HIS294
a =2.45 Å
b =1.40 Å
c =1.09 Å
d =1.58 Å
e =1.65 Å
f =1.86 Å
g =1.54 Å
h =2.33 Å

FIG. 3A

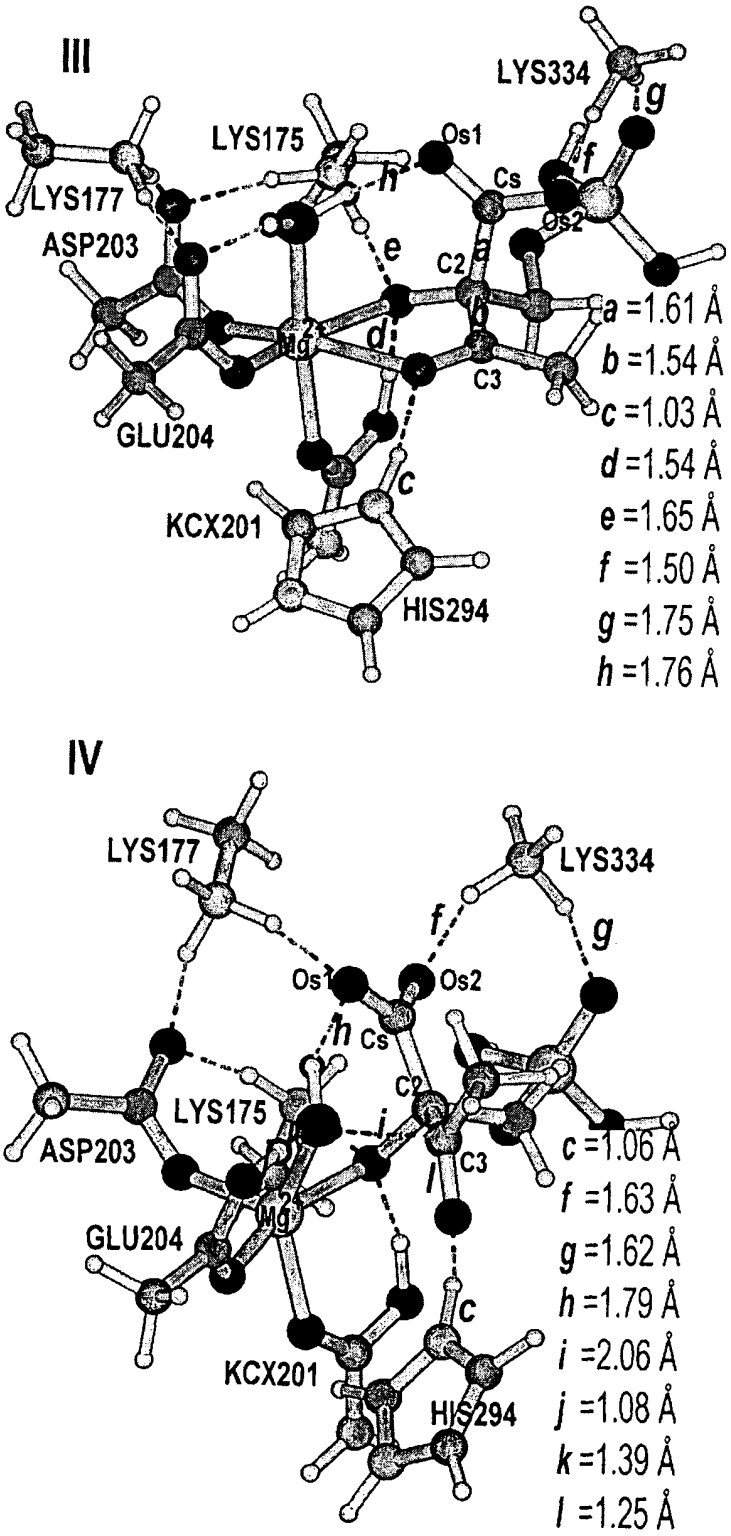

III
LYS334
LYS175
Os1
LYS177
Cs
Os2
ASP203
C2
Mg
C3
GLU204
KCX201
HIS294
a =1.61 Å
b =1.54 Å
c =1.03 Å
d =1.54 Å
e =1.65 Å
f =1.50 Å
g =1.75 Å
h =1.76 Å
IV
LYS177
LYS334
Os1
Os2
Cs
C2
LYS175
ASP203
C3
Mg
GLU204
KCX201
HIS294
c =1.06 Å
f =1.63 Å
g =1.62 Å
h =1.79 Å
i =2.06 Å
j =1.08 Å
k =1.39 Å
l =1.25 Å

FIG. 3B

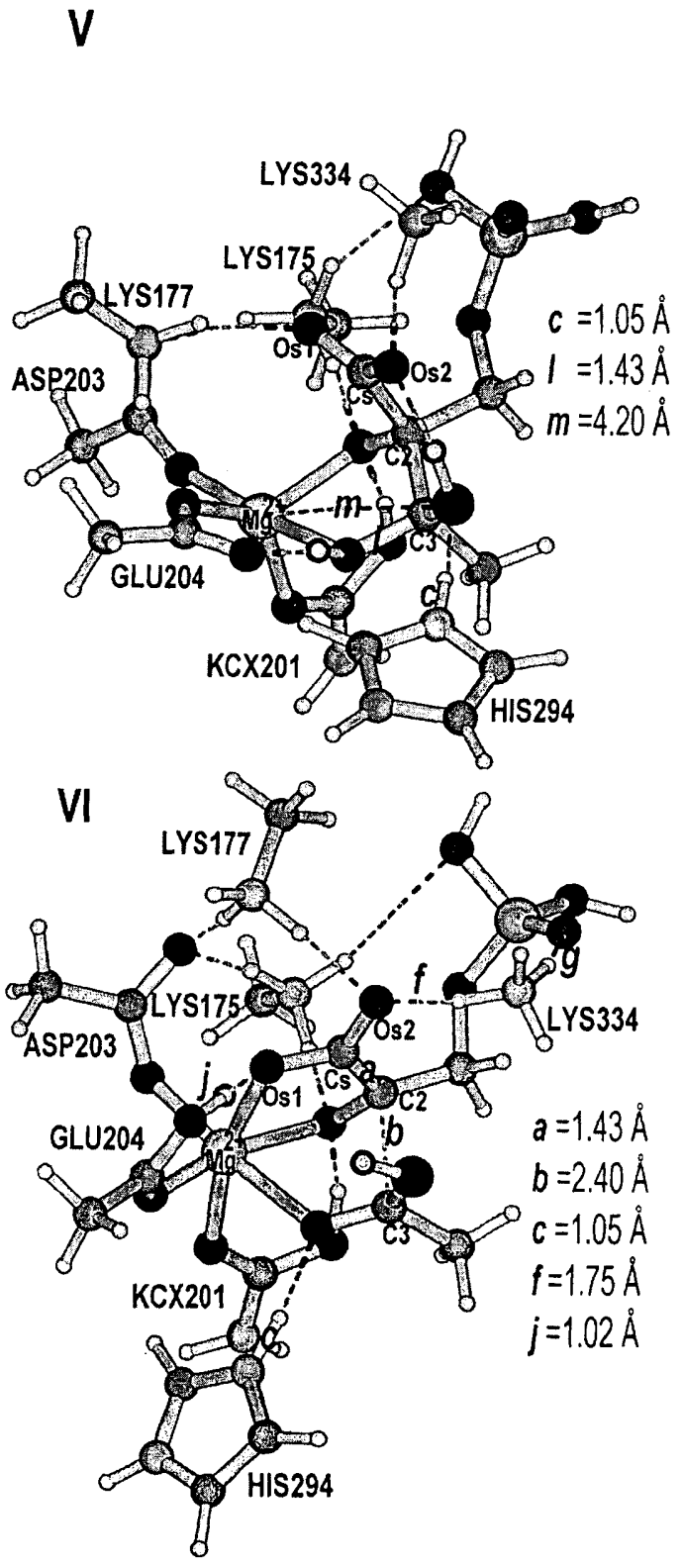

V
LYS334
LYS175
LYS177
Os1
Os2
Cs
C2
C3
ASP203
GLU204
KCX201
HIS294
Mg
m
c
l
c =1.05 Å
l =1.43 Å
m =4.20 Å
VI
LYS177
LYS175
ASP203
LYS334
GLU204
KCX201
HIS294
Os1
Os2
Cs
C2
C3
Mg
a
b
c
f
g
j
a =1.43 Å
b =2.40 Å
c =1.05 Å
f =1.75 Å
j =1.02 Å

FIG. 3C

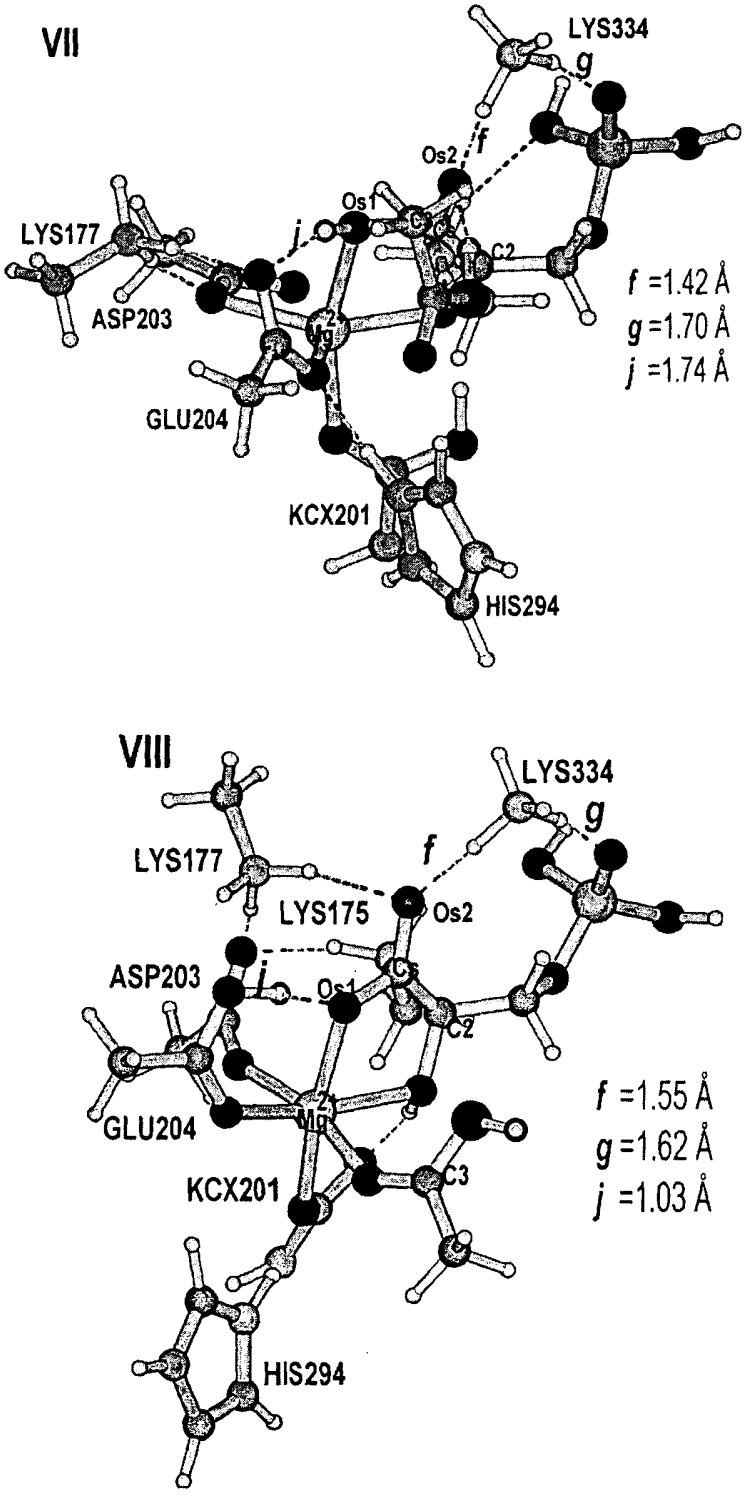

VII
LYS334
g
f
Os2
Os1
LYS177
j
C2
ASP203
f =1.42 Å
g =1.70 Å
j =1.74 Å
GLU204
KCX201
HIS294
VIII
LYS334
g
LYS177
f
LYS175
Os2
ASP203
j
Os1
C2
GLU204
KCX201
C3
f =1.55 Å
g =1.62 Å
j =1.03 Å
HIS294

FIG. 3D

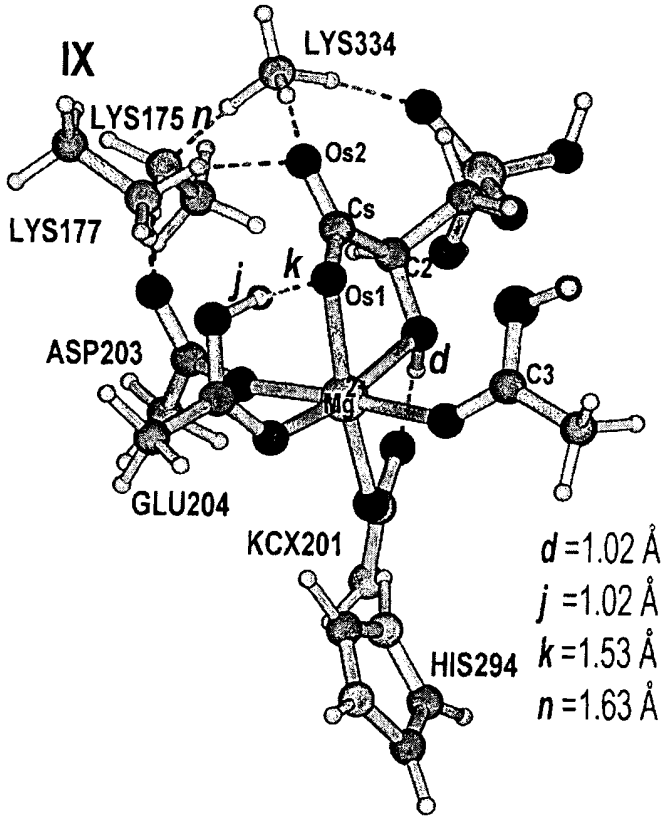

IX
LYS334
LYS175
n
Os2
Cs
LYS177
C2
j
k
Os1
ASP203
d
C3
Mg
GLU204
KCX201
HIS294
d =1.02 Å
j =1.02 Å
k =1.53 Å
n =1.63 Å

FIG. 3E

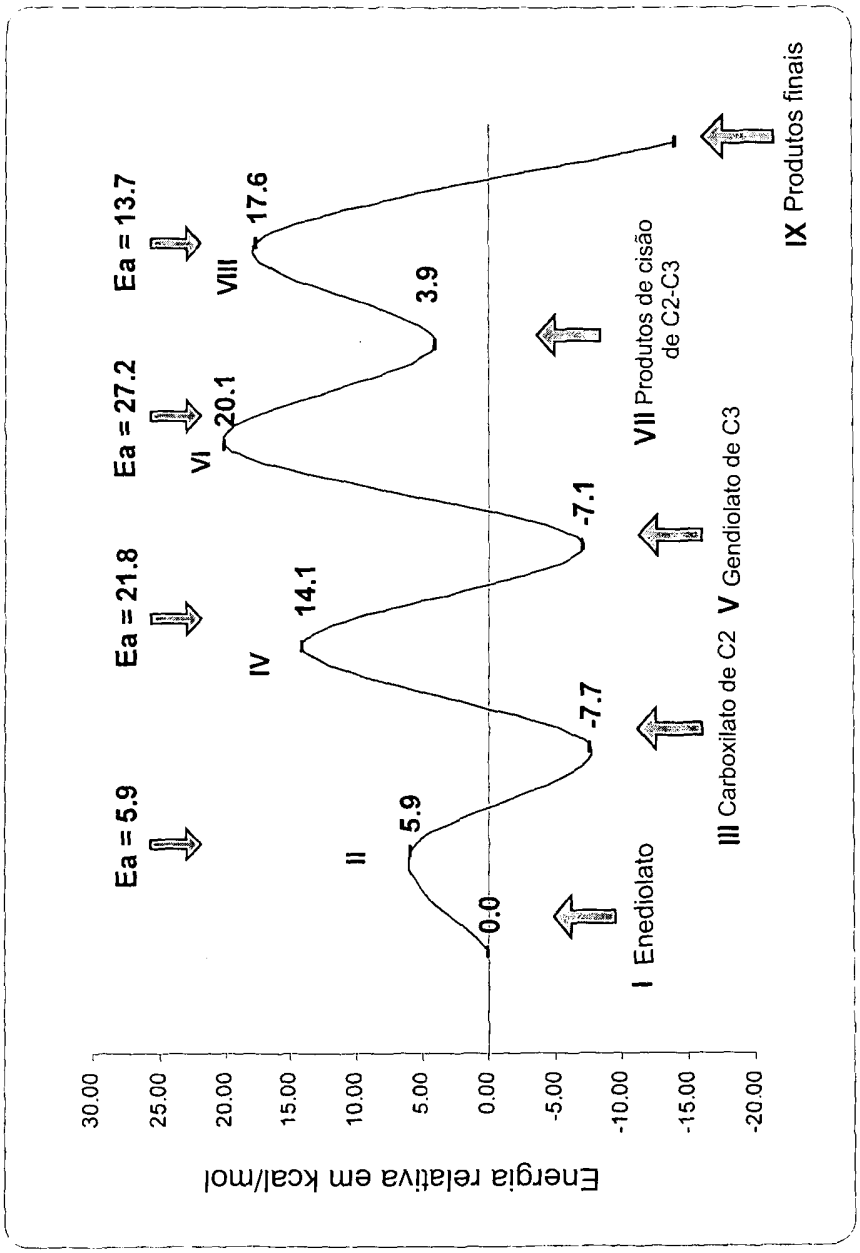

Energia relativa em kcal/mol
30.00
25.00
20.00
15.00
10.00
5.00
0.00
-5.00
-10.00
-15.00
-20.00
Ea = 5.9
Ea = 21.8
Ea = 27.2
Ea = 13.7
II
IV
VI
VIII
5.9
14.1
20.1
17.6
0.0
-7.7
-7.1
3.9
I Enediolato
III Carboxilato de C2
V Gendiolato de C3
VII Produtos de cisão de C2-C3
IX Produtos finais

FIG. 4

```
                                                17        26                  46
                                  10        20        30        40        50        60
                                  ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
protista glaucófitas              1   MSSQARTQTR A~GFK~~AGV KDYRL~TYYT PEYTPKETDI LAAFRMTPQP GVPPEECAAA  56
cns50%(9) protista rodófitas      1   MspSVpERTR IKNERYESGV IPYAKMGYWD PDYVVKDTDV LALFRVTPQP GVDPVEASAA  60
cns50%(11) eubacteria cianobactérias 1 MSYAQhKoQo KuGYp~~AGV KDYRL~TYYT PDYTPKDTDL LAAFRhTPQP GVPsEEAGAA  57
plantae antocerotófitas           1   MSPQTETKAG V~GFK~~AGV KDYRL~TYYT PDYETKDTDI LAAFRMTPQP GVPPEEAGAA  56
cns50%(28) plantae briófitas      1   MSPQTETKAG V~GFK~~AGV KDYRL~TYYT PDYQTKETDI LAAFRMTPQP GVPAEEAGAA  56
cns50%(4) plantae carófitas       1   MSPQTETKAG A~GFK~~AGV KDYRL~TYYT PDYETKETDI LAAFRMTPQA GVPPEEAGAA  56
cns50%(4) plantae clorófitas      1   MVPQTETKAG A~GFK~~AGV KDYRL~TYYT PDYVVKDTDI LAAFRMTPQP GVPPEECGAA  56
plantae coniferófitas             1   MSPKTETKAS V~GFK~~AGV KDYRL~TYYT PEYQTKDTDI LAAFRVTPQP GVPPEEAGAA  56
plantae equisetófitas             1   MSPQTETKAG V~GFK~~AGV KDYRL~TYFT PDYETKDTDI LAAFRMTPQP GVPPEEAGAA  56
plantae gnetófitas                1   MSPKTETKAS V~GFQ~~AGV KDYRL~TYYT PEYQTKDTDI LAAFRVTPQP GVPPEEAGAA  56
cns50%(134) plantae magnoliófitas 1   MSPQTETKAS V~GFK~~AGV KDYKL~TYYT PDYETKDTDI LAAFRVTPQP GVPPEEAGAA  56
cns50%(2) plantae pinófitas       1   MSPKTETKAS V~GFK~~AGV KDYRL~TYYT PEYQTKDTDI LAAFR       GVPAEEAGAA  56
cns50%(2) plantae pteridófitas    1   MSPQTETKTG I~GFK~~AGV KDYRL~TYYT PDYETKDTDI LAAFR       GVPAEEAGAA  56
                                                76                  96                  116
                                      70        80        90        100                 120
                                  ....|....| ....|....| ....|....| ....|....| ....|      ....|....|
protista glaucófitas              57  VAAESSTGTW TTVWTDGLTS LDRYKGRSYG FEPVPGEENQ YICYV       LFEEGSVTNM
116
cns50%(9) protista rodófitas      61  VAGESSTATW TVVWTDLLTA CDLYRAKAYK VDuVPNoSDQ YFAYI       LFEEGSIANL  120
cns50%(11) eubacteria cianobactérias 58 VAAESSTGTW TTVWTDLLTD LDRYKGRCYD IEPVPGEDNQ YaAFI       LFEEGSVTNV  117
plantae antocerotófitas           57  VAAESSTGTW TTVWTDGLTS LDRYKGRCYD IEPVAGEENQ YIAYVAYPLD LFEEGSVTNM  116
cns50%(28) plantae briófitas      57  VAAESSTGTW TTVWTDGLTS LDRYKGRCYD IEAVPGEENQ YIAYVAYPLD LFEEGSVTNL  116
cns50%(4) plantae carófitas       57  VAAESSTGTW TTVWTDGLTS LDRYKGRCYD IEPVAGEENQ YIAYVAYPLD LFEEGSVTNL  116
cns50%(4) plantae clorófitas      57  VAAESSTGTW TTVWTDGLTS LDRYKGRCYD IEPVPGEDNQ YIAYVAYPID LFEEGSVTNL  116
plantae coniferófitas             57  VAAESSTGTW TTVWTDGLTS LDRYKGRCYD IEPVPGEETQ FIAYVAYPLD LFEEGSVTNL  116
plantae equisetófitas             57  VAAESSTGTW TTVWTDGLTS LDRYKGRCYN IEPVAGEDNQ FIAYVAYPLD LFEEGSVTNL  116
plantae equisetófitas             57  VAAESSTGTW TTVWTDGLTS LDRYKGRCYD LEPVPGEDNQ YIAYVAYPLD LFEEGSVTNM  116
cns50%(134) plantae magnoliófitas 57  VAAESSTGTW TTVWTDGLTS LDRYKGRCYH IEPVAGEENQ FIAYVAYPLD LFEEGSVTNM  116
cns50%(2) plantae pinófitas       57  VAAESSTGTW TTVWTDGLTS LDRYKGRCYD IEPVPGEENQ FIAYVAYPLD LFEEGSVTNL  116
cns50%(2) plantae pteridófitas    57  VAAESSTGTW TTVWTDGLTS LDRYKGRCYD IEPVAGEENQ YIAYVAYPLD LFEEGSVTNM  116
```

FIG. 5A

```
                                                136                 156                 176
                                  130           140           150           160           170           180
                                  ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
protista glaucófitas              117 LTSIVGNVFG FKALRALRLE DLRIPVGYSK TFQGPPHGIT VERDKLNKYG RALLGCTIKP 176
cns50%(9) protista rodófitas      121 TASIIGNVFG FKAVKALRLE DMRLPVAYLK TFQGPATGlI VERERMDKFG RPFLGATVKP 180
cns50%(11) eubacteria cianobactérias 118 LTSIVGNVFG FKALRALRLE DIRFPVAaIK TFQGPPHGIQ VERDKLNKYG RPLLGCTIKP 177
plantae antocerotófitas           117 FTSIVGNVFG FKALRALRLE DLRIPPAYSK TFQGPPHGIQ VERDKLNKYG RPLLGCTIKP 176
cns50%(28) plantae briófitas      117 FTSIVGNVFG FKALRALRLE DLRIPPAYSK TFQGPPHGIQ VERDKLNKYG RPLLGCTIKP 176
cns50%(4) plantae carófitas       117 FTSIVGNVFG FKALRALRLE DLRIPPAYSK TFQGPPHGIQ VERDKINKYG RPLLGCTIKP 176
cns50%(4) plantae clorófitas      117 FTSIVGNVFG FKALRALRLE DLRIPPAYVK TFQGPPHGIQ VERDKLNKYG RGLLGCTIKP 176
plantae coniferófitas             117 FTSIVGNVFG FKALRALRLE DLRIPPSYSK TFQGPPHGIQ VERDKLNKYG RPLLGCTIKP 176
plantae equisetófitas             117 FTSIVGNVFG FKALRALRLE DLRIPPAYSK TFIGPPHGIQ VERDKLNKYG RPLLGCTIKP 176
plantae gnetófitas                117 FTSIVGNVFG FKALRALRLE DLRIPTSYIK TFQGPPHGIQ VERDKLNKYG RPLLGCTIKP 176
cns50%(134) plantae magnoliófitas 117 FTSIVGNVFG FKALRALRLE DLRIPsAYVK TFQGPPHGIQ VERDKLNKYG RPLLGCTIKP 176
cns50%(2) plantae pinófitas       117 FTSIVGNVFG FKALRALRLE DLRIPPAYSK TFQGPPHGIQ VERDKLNKYG RPLLGCTIKP 176
cns50%(2) plantae pteridófitas    117 LTSIVGNVFG FKALRALRLE DLRILPAYSK TFIGPPHGIQ VERDKLNKYG RPLLGCTIKP 176
                                                196                 216                 236
                                  190           200           210           220           230           240
                                  ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
protista glaucófitas              177 KLGLSAKNYG RAVYECLRGG LDFTKDDENV NSQPFMRWRD RFLYVMDAIK KSQAETGEIK 236
cns50%(9) protista rodófitas      181 KLGLSGKNYG RVVYEGLKGG LDFLKDDENI NSQPFMRW+E RYLYSMEGVN RApAAoGElK 240
cns50%(11) eubacteria cianobactérias 178 KLGLSAKNYG RAVYECLRGG LDFTKDDENI NSQPFQRWRD RFLFVAEAIc KAQAETGEIK 237
plantae antocerotófitas           177 KLGLSAKNYG RAVYECLRGG LDFTKDDENV NSQPFMRWRD RFLFVAEAIF KSQAETGEIK 236
cns50%(28) plantae briófitas      177 KLGLSAKNYG RAVYECLRGG LDFTKDDENV NSQPFMRWRD RFLFVAEAIY KSQAETGEIK 236
cns50%(4) plantae carófitas       177 KLGLSAKNYG RAVYECLRGG LDFTKDDENV NSQPFMRWRD RFLFVAEAIF KAQAETGEIK 236
cns50%(4) plantae clorófitas      177 KLGLSAKNYG RAVYECLRGG LDFTKDDENV NSQPFMRWRD RFLFVAEAIY KAQAETGEVK 236
plantae coniferófitas             177 KLGLSAKNYG RAVYECLRGG LDFTKDDENV NSQPFMRWRD RFVFCAEALN KAQAETGEIK 236
plantae equisetófitas             177 KLGLSAKNYG RAVYECLRGG LDFTKDDENV NSQPFMRWRD RFLFVAEALF KSQAETGEIK 236
plantae equisetófitas             177 KLGLSAKNYG RAVYECLRGG LDFTKDDENV NSQPFMRWRD RFVFCAEAIY KAQAETGEIK 236
cns50%(134) plantae magnoliófitas 177 KLGLSAKNYG RAVYECLRGG LDFTKDDENV NSQPFMRWRD RFLFCAEAIY KAQAETGEIK 236
cns50%(2) plantae pinófitas       177 KLGLSAKNYG RAVYECLRGG LDFTKDDENV NSQPFMRWRD RFVFCAEAIN KAQAETGEIK 236
cns50%(2) plantae pteridófitas    177 KLGLCAKNYG RAVYECLRGG LDFTKDDENV NSQPFMRWRD RFLFVAEALF KAQAETGEIK 236
```

FIG. 5B

```
                                        256                 276                 296
                          250           260       270       280       290       300
                          ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
protista glaucófitas      237 GHYLNATAPT TEEMIKRAEF AAELDAPIIM HDYITAGFTS NTTLARWCRD NGPLLHIHRA 296
cns50%(9) protista rodófitas 241 GHYLNVTAAT ME-MYERAEF AKELGSVIlM ID~LVIGYTA IQTMAIWARK NDMILHLHRA 298
cns50%(11) eubacteria cianobactérias 238 GHYLNVTAsT CEEMhKRAEF AKELGpPIIM HDFLTGGFTA NTTLAKWCRD NGlLLHIHRA 297
plantae antocerotófitas   237 GHYLNATAGT CEEMMKRAHF ARELGMPIVM HDYLTGGFTA NTTLARYCRD NGLLLHIHRA 296
cns50%(28) plantae briófitas 237 GHYLNATAGT CEEMLKRAQF ARELGMPIVM HDYLTGGFTA NTSLAHYCRD NGLLLHIHRA 296
cns50%(4) plantae carófitas 237 GHYLNATAGT CEEMLKRAAY AKELGVPIIM HDYLTGGFTA NTSLAHYCRD NGLLLHIHRA 296
cns50%(4) plantae clorófitas 237 GHYLNATAAT CEEMLKRAVC AKELGVPIIM HDYLTGGFTA NTSLAsYCRD NGLLLHIHRA 296
plantae coniferófitas      237 GHYLNATAGT CEEMMKRAIF ARELGVPIVM HDYLTGGFTA NTSLAHYCRD NGLLLHIHRA 296
plantae equisetófitas      237 GHYLNATAGT CEEMLKRAVF ARELGAPIVM HDYLTGGFTA NTSLAFYCRD NGLLLHIHRA 296
plantae gnetófitas         237 GHYLNATAGT CEEMIKRAVF ARELGVPIVM HDYLTGGFTA NTTLAHYCRD NGLLLHIHRA 296
cns50%(134) plantae magnoliófitas 237 GHYLNATAGT CEEMlKRAVF ARELGVPIVM HDYLTGGFTA NTSLAHYCRD NGLLLHIHRA 296
cns50%(2) plantae pinófitas 237 GHYLNATAGT CEEMMKRAIF ARELGVPIVM HDYLTGGFTA NTSLAHYCRD NGLLLHIHRA 296
cns50%(2) plantae pteridófitas 237 GHYLNATAAT CEEMIKRAIF ARELGAPIVM HDYLTGGFTA NTSLAFYCRD NGLLLHIHRA 296
                                        316                 336                 356
                          310           320       330       340       350       360
                          ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
protista glaucófitas      297 MHAVIDRQKN HGIHFRVLAK TLRMSGGDHL HSGTVVGKLE GDRAGTLGFV DLMRDDHIEQ 356
cns50%(9) protista rodófitas 299 GNSTYSRQKs HGMNFRVICK WMRMAGVDHI HAGTVVGKLE GDPlMI+GFY NTLLLoHL-l 357
cns50%(11) eubacteria cianobactérias 298 MHAVIDRQKN HGIHFRVLAK CLRLSGGDHL HTGTVVGKLE GERuITMGFV DLLRE-aVEc 356
plantae antocerotófitas    297 MHAVIDRQRN HGIHFRVLAK ALRMSGGDHI HSGTVVGKLE GEREVTLGFV DLLRDDYIEK 356
cns50%(28) plantae briófitas 297 MHAVIDRQKN HGMHFRVLAK ALRLSGGDHI HAGTVVGKLE GERQVTLGFV DLLRDDYIEK 356
cns50%(4) plantae carófitas 297 MHAVIDRQKN HGIHFRVLAK ALRMSGGDHI HSGTVVGKLE GEREVTLGFV DLLRDDYIEK 356
cns50%(4) plantae clorófitas 297 MHAVIDRQRN HGIHFRVLAK ALRLSGGDHL HSGTVVGKLE GEREVTLGFV DLMRDDYIEK 356
plantae coniferófitas       297 MHAVIDRQRN HGMHFRVLAK ALRMSGGDHI HAGTVVGKLE GERDVTLGFV DLLRDDFIEK 356
plantae equisetófitas       297 MHAVIDRQKN HGIHFRVLAK ALRMSGGDHI HTGTVVGKLE GERDLTLGFV DLLRDDFIEK 356
plantae equisetófitas       297 MHAVIDRQKN HGMHFRVLAK ALRMSGGDHI HAGTVVGKLE GEREITLGFV DLLRDDFIEK 356
cns50%(134) plantae magnoliófitas 297 MHAVIDRQKN HGMHFRVLAK ALRMSGGDHI HAGTVVGKLE GEREITLGFV DLLRDDFIEK 356
cns50%(2) plantae pinófitas 297 MHAVIDRQKN HGMHFRVLAK ALRMSGGDHI HAGTVVGKLE GERDVTLGFV DLLRDDFIEK 356
cns50%(2) plantae pteridófitas 297 MHAVIDRQKN HGMHFRVLAK ALRMSGGDHI HAGTVVGKLE GEREVTLGFV DLLRDDYIEK 356
```

FIG. 5C

```
                                                376                 396                 416
                                370         380         390         400         410         420
                                ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
protista glaucófitas            357 DRSRGIFFTQ DWASMPGVMP VASGGIHIWH MPALVDIFGD DSCLQFGGGT LGHPWGNAPG 416
cns50%(9) protista rodófitas    358 NLPQGIFFEQ DWASLRKVTP VASGGIHCGQ MHQLLDYLGD DVVLQFGGGT IGHPDGIQAG 417
cns50%(11) eubacteria cianobactérias 357 DRSRGIFFTQ DWASMPGVMA VASGGIHVWH MPALVEIFGD DSVLQFGGGT LGHPWGNAPG 416
plantae antocerotófitas          357 DRSRGIYFTQ DWVSMPGVLP VASGGIHVWH MSALTEIFGD DSVLQFGGGT LGHPWGNAPG 416
cns50%(28) plantae briófitas     357 DRSRGIYFTQ DWVSLPGVLP VASGGIHVWH MPALTEIFGD DSVLQFGGGT LGHPWGNAPG 416
cns50%(4) plantae carófitas      357 DRSRGIYFTQ DWVSMPGVLP VASGGIHVWH MPALTEIFGD DSVLQFGGGT LGHPWGNAPG 416
cns50%(4) plantae clorófitas     357 DRSRGIYFTQ DWCSMPGVMP VASGGIHVWH MPALVEIFGD DACLQFGGGT LGHPWGNAPG 416
plantae coniferófitas            357 DRSRGIYFTQ DWVSMPGVLP VASGGIHVWH MPALTEIFGD DSVLQFGGGT LGHPWGNAPG 416
plantae equisetófitas            357 DRSRGIYFTQ DWVSMPGVLP VASGGIHVWH MPALTEIFGD DSVLQFGGGT LGHPWGNAPG 416
plantae gnetófitas               357 DRSRGIYFTQ DWVSMPGVMP VASGGIHVWH MPALTDIFGD DAVLQFGGGT LGHPWGNAPG 416
cns50%(134) plantae magnoliófitas 357 DRSRGIYFTQ DWVSLPGVLP VASGGIHVWH MPALTEIFGD DSVLQFGGGT LGHPWGNAPG 416
cns50%(2) plantae pinófitas      357 DRSRGIYFTQ DWVSMPGVLP VASGGIHVWH MPALTEIFGD DSVLQFGGGT LGHPWGNAPG 416
cns50%(2) plantae pteridófitas   357 DRSRGIYFTQ DWVSMPGVFP VASGGIHVWH MPALTEIFGD DSVLQFGGGT LGHPWGNAPG 416

                                                436                 456
                                430         440         450         460         470         480
                                ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
protista glaucófitas            417 AVANRVALEA CVQARNEGRN LAREGNEIIR EAARFSPELA AACEVWKEIK FEFETIDTI  475
cns50%(9) protista rodófitas    418 ATANRVALES MVIARNEGRD YVAEGPQILR DAAKTCGPLQ TALDLWKDIT FNYTSTDTAD 477
cns50%(11) eubacteria cianobactérias 417 ATANRVALEA CVQARNEGRc LtREGsDIIR EAAKWSPELA lACELWKEIK FEFEAMDTL  475
plantae antocerotófitas          417 AVANRVALEA CVQARNEGRD LAREGNDIIR EASKWSPELA AACEVWKEIK FVFETIDTL  475
cns50%(28) plantae briófitas     417 AVANRVALEA CVQARNEGRD LAREGNEVIR EAAKWSPELA AACEVWKEIK FEFETIDTV  475
cns50%(4) plantae carófitas      417 AVANRVALEA CVQARNEGRD LAREGNDVIR EACKWSPELA AACEVWKEIK FEFDTIDTL  475
cns50%(4) plantae clorófitas     417 AAANRVALEA CTQARNEGRD LAREGGDVIR SACKWSPELA AACEVWKEIK FEFDTIDKL  475
plantae coniferófitas            417 AVANRVALEA CVQARNEGRD LAREGNEVIR EACKWSPELA AACEIWKEIK FEFDVIDRL  475
plantae equisetófitas            417 AVANRVAVEA CVQARNEGRD LATEGNDIIR EAAKWSPELA AACEVWKEIK FEYEAMDTL  475
plantae equisetófitas            417 AVANRVALEA CVQARNEGRD LAREGNEVIR EAAKWSPELA AACEVWKEIK FEFESVDTL  475
cns50%(134) plantae magnoliófitas 417 AVANRVALEA CVQARNEGRD LAREGNEIIR EASKWSPELA AACEVWKEIK FEFEAMDTL  475
cns50%(2) plantae pinófitas      417 AVANRVALEA CVQARNEGRD LAREGNEVIR EACKWSPELA AACEIWKEIK FEFDAIDRL  475
cns50%(2) plantae pteridófitas   417 AVANRVALEA CVQARNEGRD LAREGNEIIR EASKWSPELA AACEVWKEIK FEFETIDTV  475
```

FIG. 5D

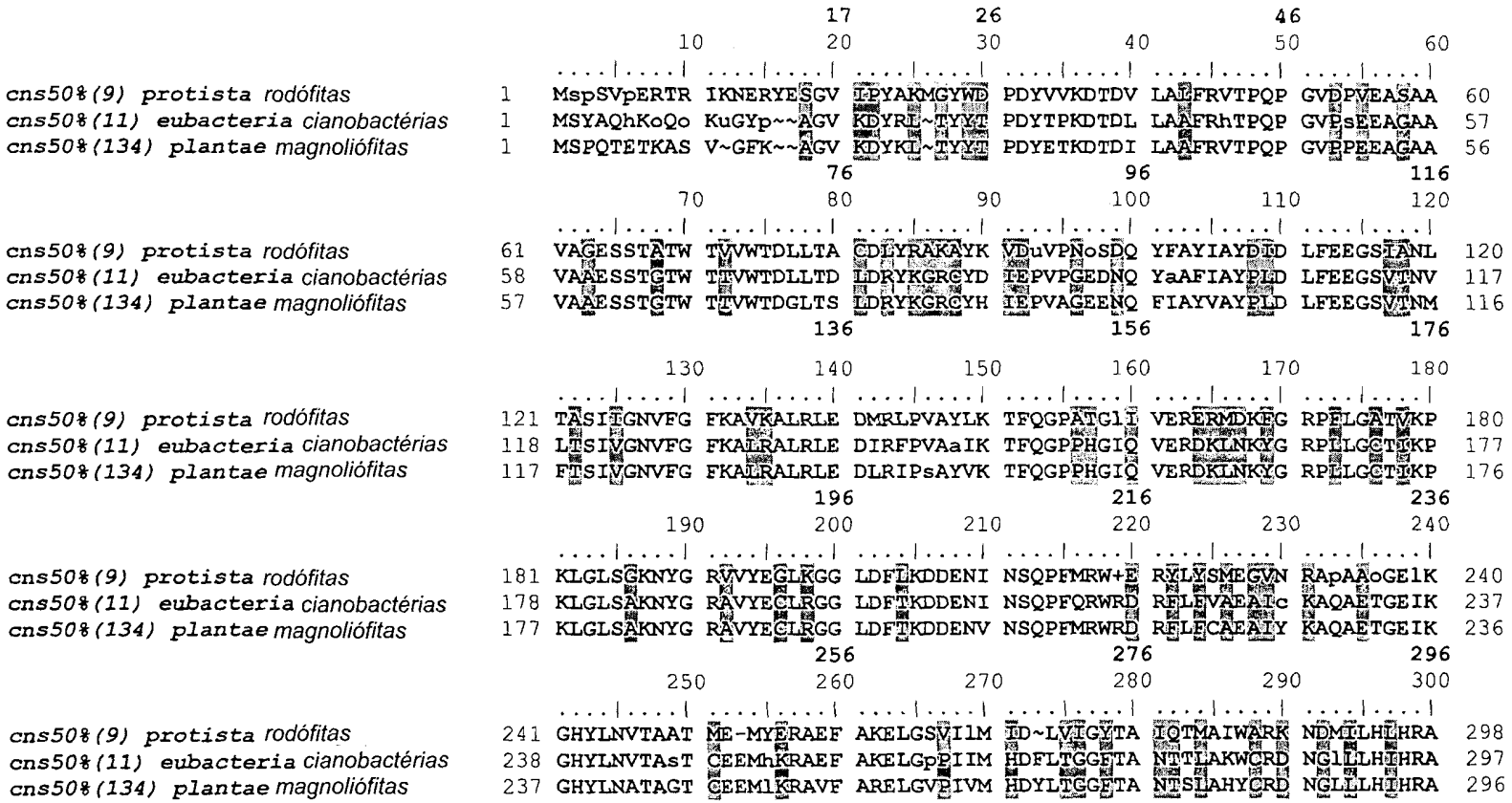

cns50%(9) protista rodófitas
cns50%(11) eubacteria cianobactérias
cns50%(134) plantae magnoliófitas

FIG. 6A

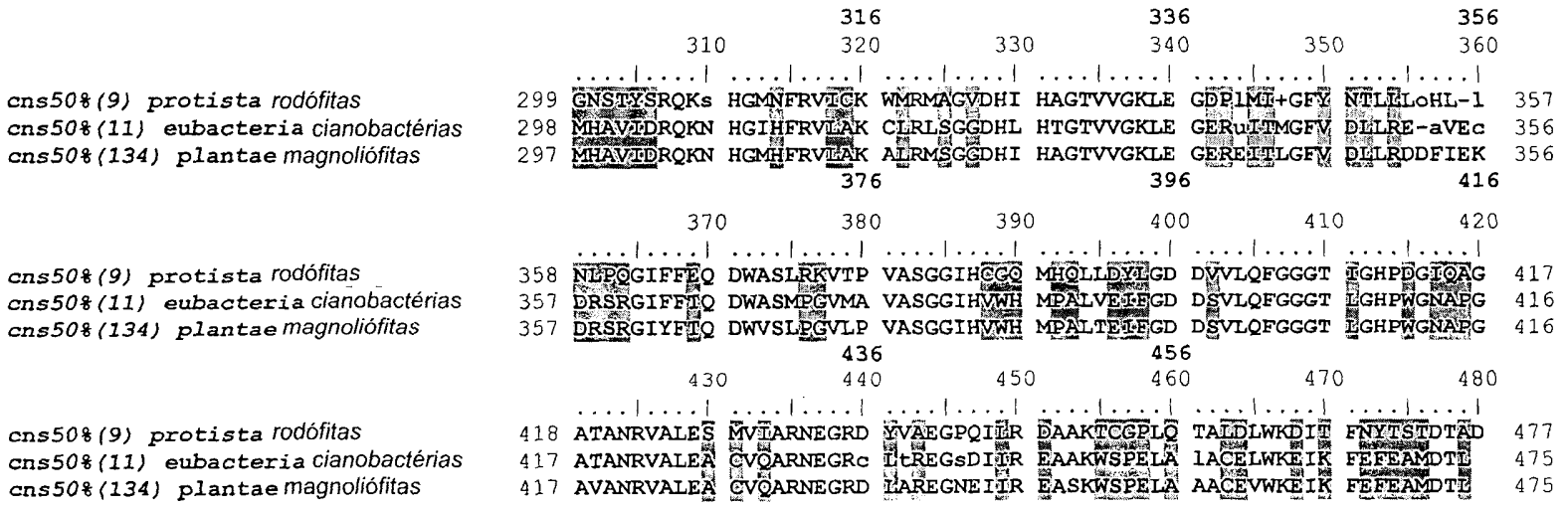

316
336
356
310
320
330
340
350
360
cns50%(9) protista rodófitas 299 GNSTYSRQKs HGMNFRVTCK WMRMAGVDHI HAGTVVGKLE GDPLMI+GFY NTLLLoHL-1 357
cns50%(11) eubacteria cianobactérias 298 MHAVIDRQKN HGIHFRVLAK CLRLSGGDHL HTGTVVGKLE GERuITMGFV DLLRE-aVEc 356
cns50%(134) plantae magnoliófitas 297 MHAVIDRQKN HGMHFRVLAK ALRMSGGDHI HAGTVVGKLE GEREITLGFV DLLRDDFIEK 356
376
396
416
370
380
390
400
410
420
cns50%(9) protista rodófitas 358 NLPQGIFFIQ DWASLRKVTP VASGGIHCGQ MHQLLDYLGD DVVLQFGGGT IGHPDGIQAG 417
cns50%(11) eubacteria cianobactérias 357 DRSRGIFFIQ DWASMPGVMA VASGGIHVWH MPALVEIFGD DSVLQFGGGT LGHPWGNAPG 416
cns50%(134) plantae magnoliófitas 357 DRSRGIYFIQ DWVSLPGVLP VASGGIHVWH MPALTEIFGD DSVLQFGGGT LGHPWGNAPG 416
436
456
430
440
450
460
470
480
cns50%(9) protista rodófitas 418 ATANRVALES MVIARNEGRD YVAEGPQIIR DAAKTCGPLQ TALDLWKDIT FNYTSTDTAD 477
cns50%(11) eubacteria cianobactérias 417 ATANRVALEA CVQARNEGRc LtREGsDIIR EAAKWSPELA lACELWKEIK FEFEAMDTL 475
cns50%(134) plantae magnoliófitas 417 AVANRVALEA CVQARNEGRD LAREGNEIIR EASKWSPELA AACEVWKEIK FEFEAMDTL 475

FIG. 6B

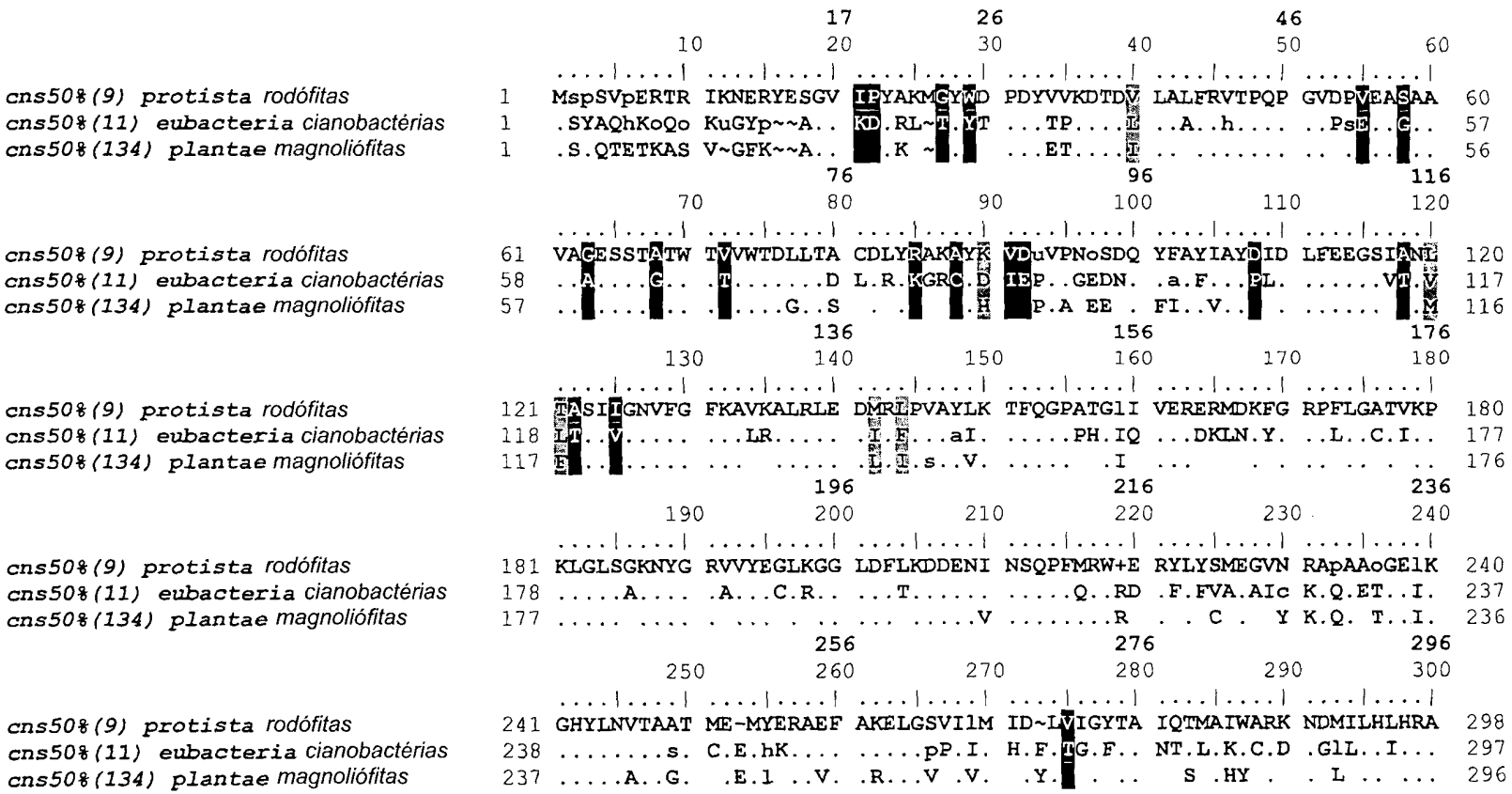

cns50%(9) protista rodófitas
cns50%(11) eubacteria cianobactérias
cns50%(134) plantae magnoliófitas

FIG. 7A

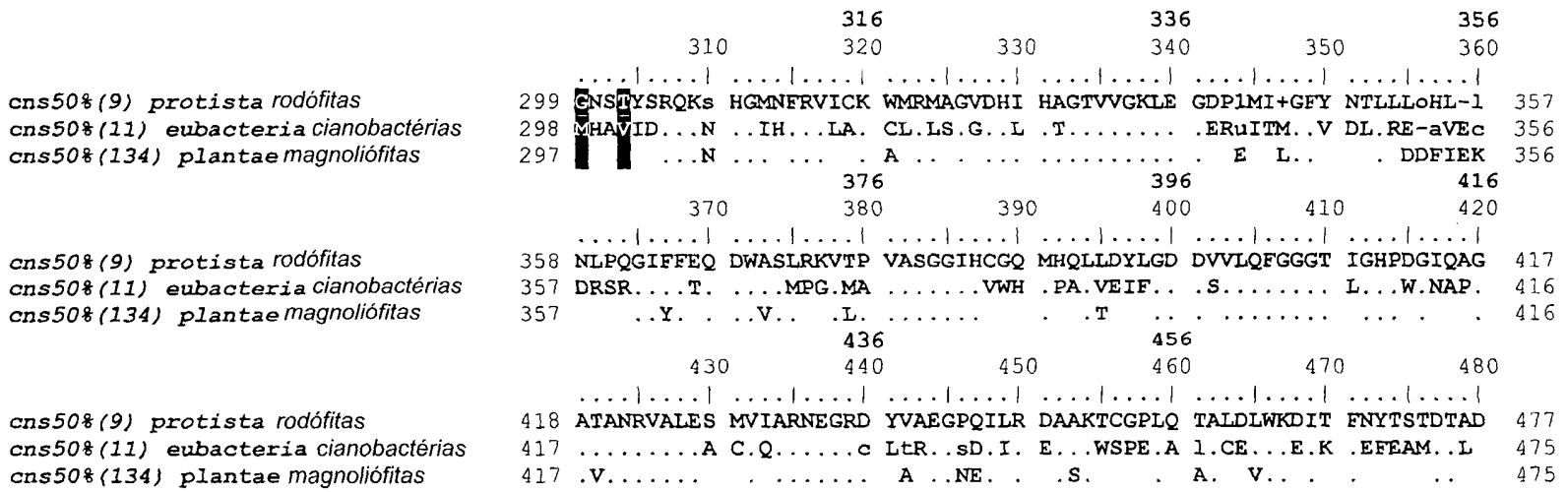

316 336 356
310 320 330 340 350 360
cns50%(9) protista rodófitas 299 GNSTYSRQKs HGMNFRVICK WMRMAGVDHI HAGTVVGKLE GDPlMI+GFY NTLLLoHL-l 357
cns50%(11) eubacteria cianobactérias 298 MHAVID...N ..IH...LA. CL.LS.G..L .T..... .ERuITM..V DL.RE-aVEc 356
cns50%(134) plantae magnoliófitas 297 ...N ... . A .. . . E L.. . DDFIEK 356
376 396 416
370 380 390 400 410 420
cns50%(9) protista rodófitas 358 NLPQGIFFEQ DWASLRKVTP VASGGIHCGQ MHQLLDYLGD DVVLQFGGGT IGHPDGIQAG 417
cns50%(11) eubacteria cianobactérias 357 DRSR....T. ....MPG.MA .....VWH .PA.VEIF.. .S..... L...W.NAP. 416
cns50%(134) plantae magnoliófitas 357 ..Y. . ..V.. .L. ..... . .T 416
436 456
430 440 450 460 470 480
cns50%(9) protista rodófitas 418 ATANRVALES MVIARNEGRD YVAEGPQILR DAAKTCGPLQ TALDLWKDIT FNYTSTDTAD 477
cns50%(11) eubacteria cianobactérias 417 .....A C.Q.....c LtR..sD.I. E...WSPE.A l.CE...E.K .EFEAM..L 475
cns50%(134) plantae magnoliófitas 417 .V..... . ..... A ..NE. . .S. . A. V.. 475

FIG. 7B

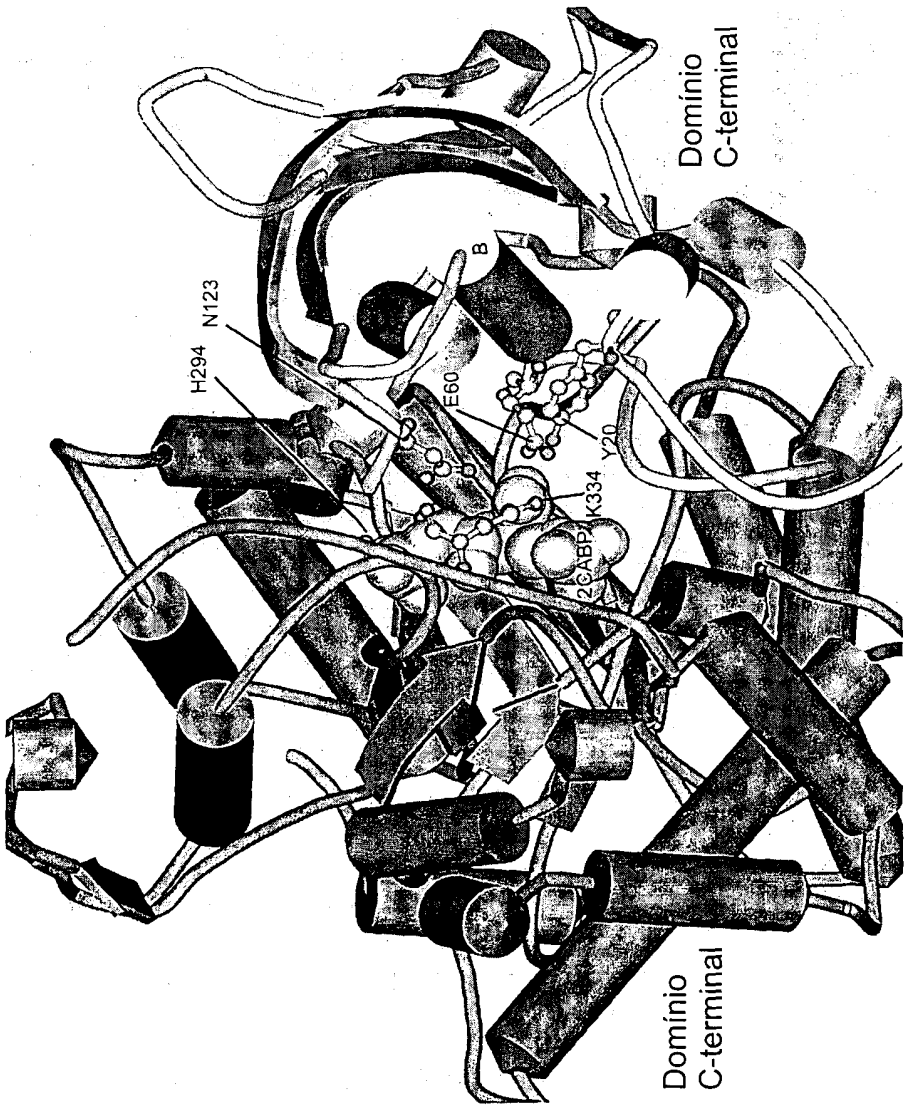

Domínio
C-terminal
B
H294
N123
E60
Y20
K334
2CABP
Domínio
C-terminal

FIG. 8

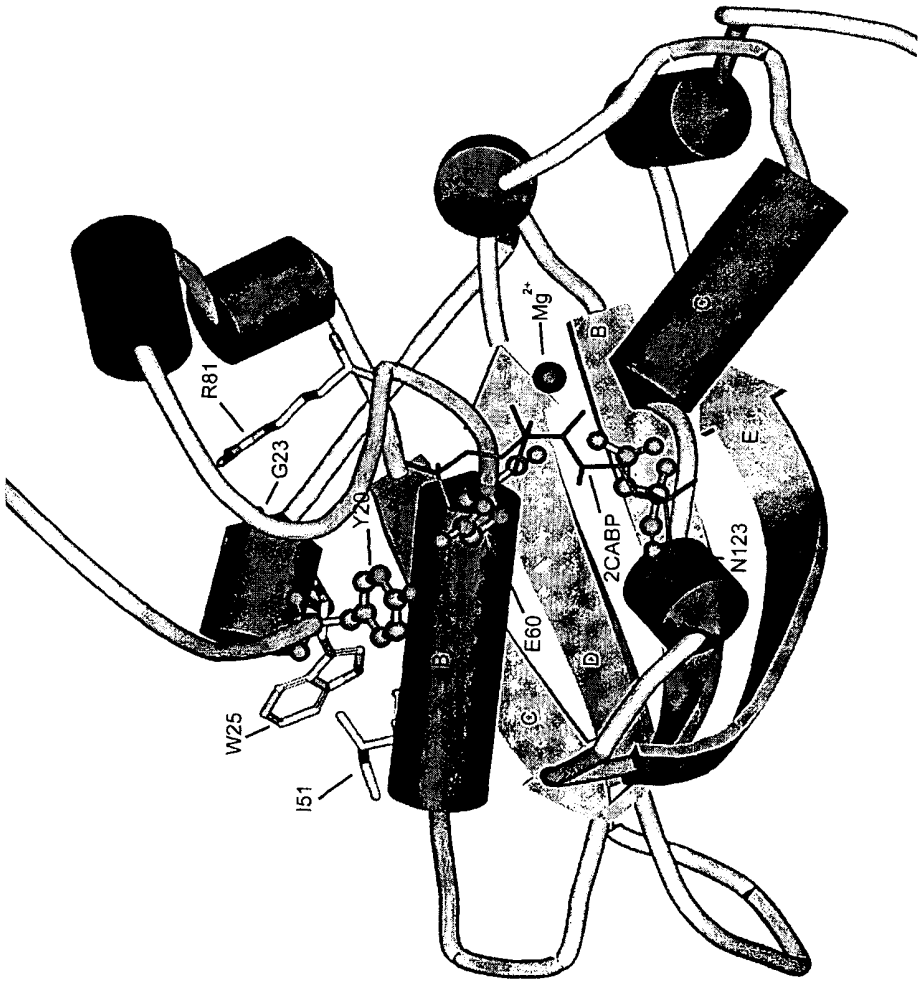


FIG. 9

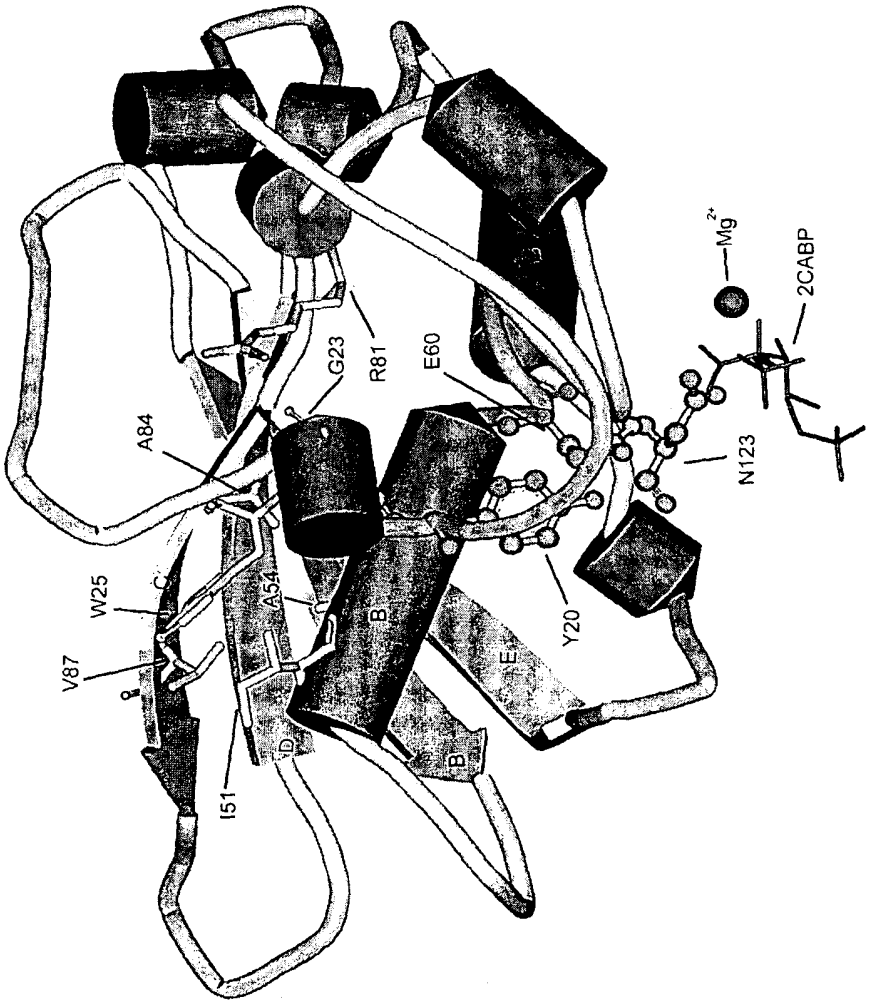


FIG. 10

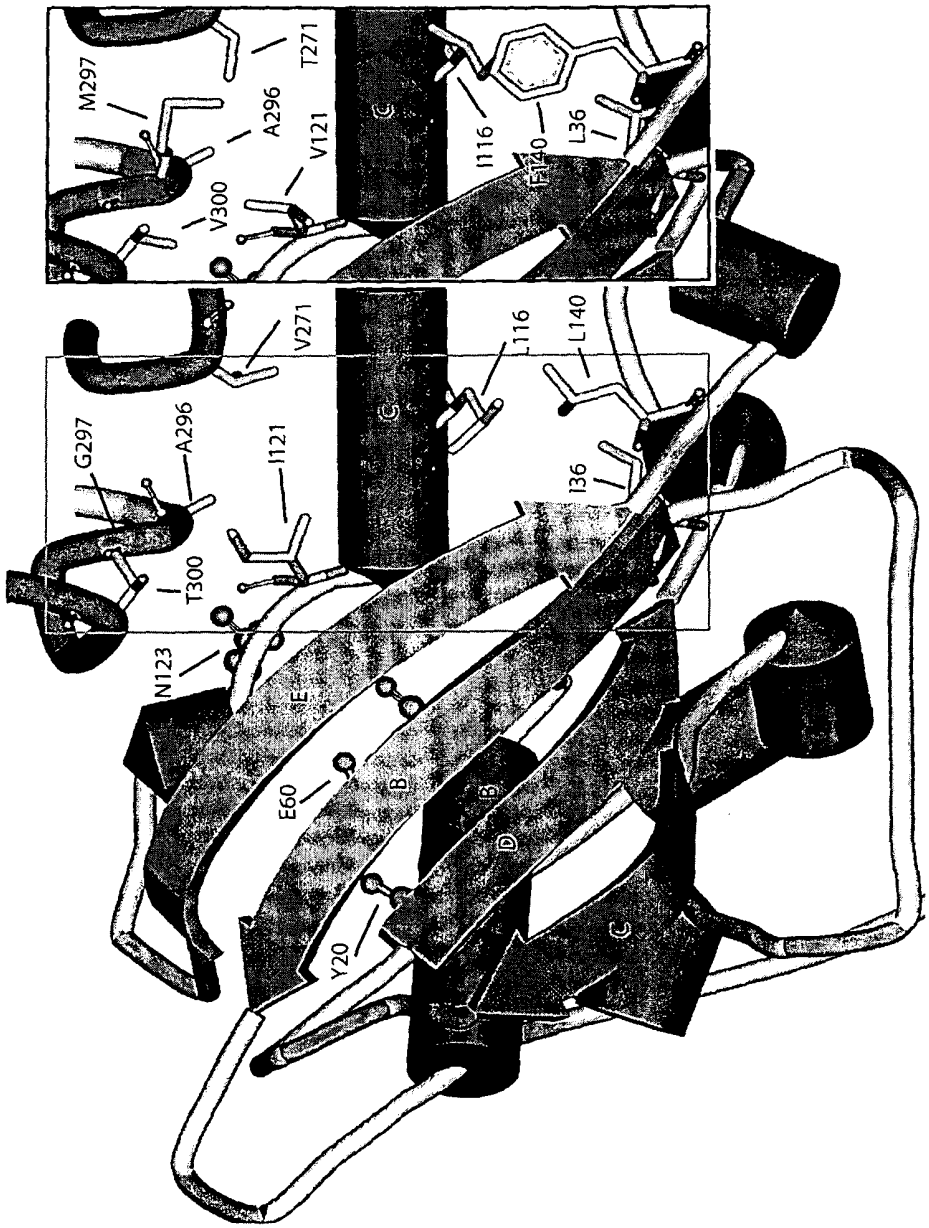

M297
A296
T271
V300
V121
I116
L36
L140
G297
A296
V271
I121
L116
L140
T300
I36
N123
E60
Y20
E
B
B
D
C
C

FIG. 11

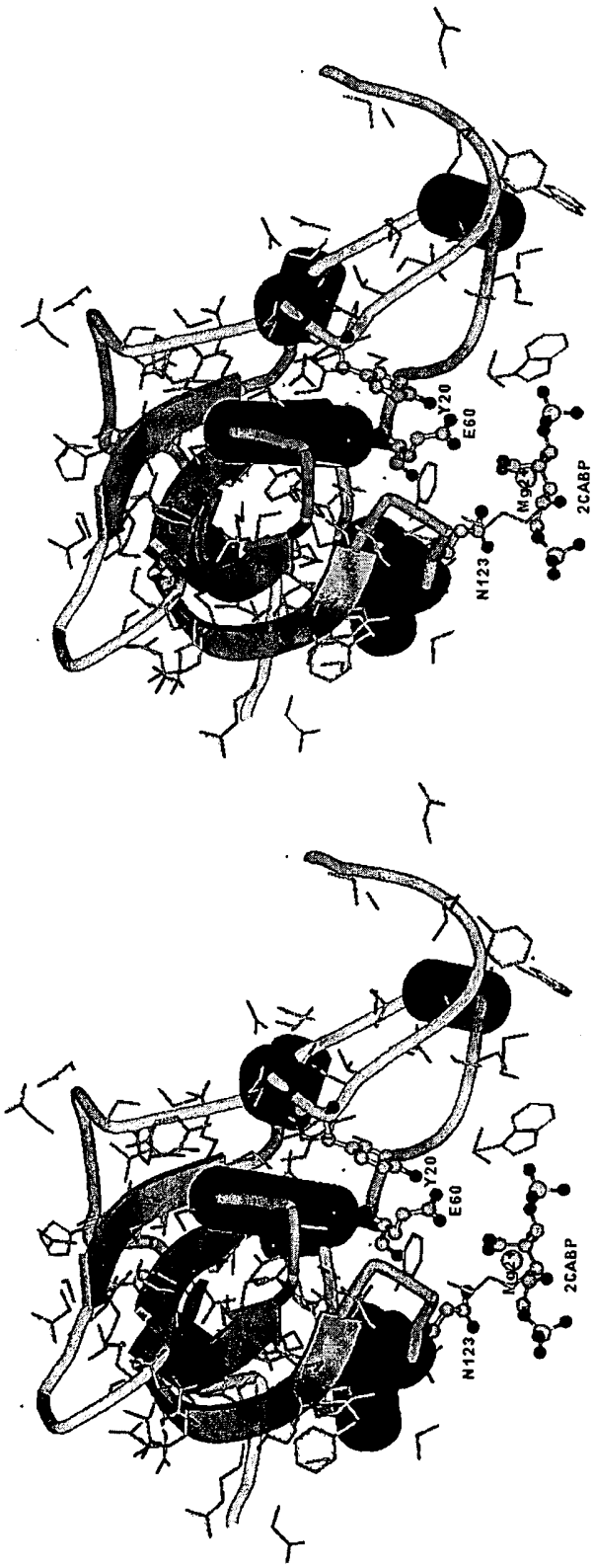


FIG. 12

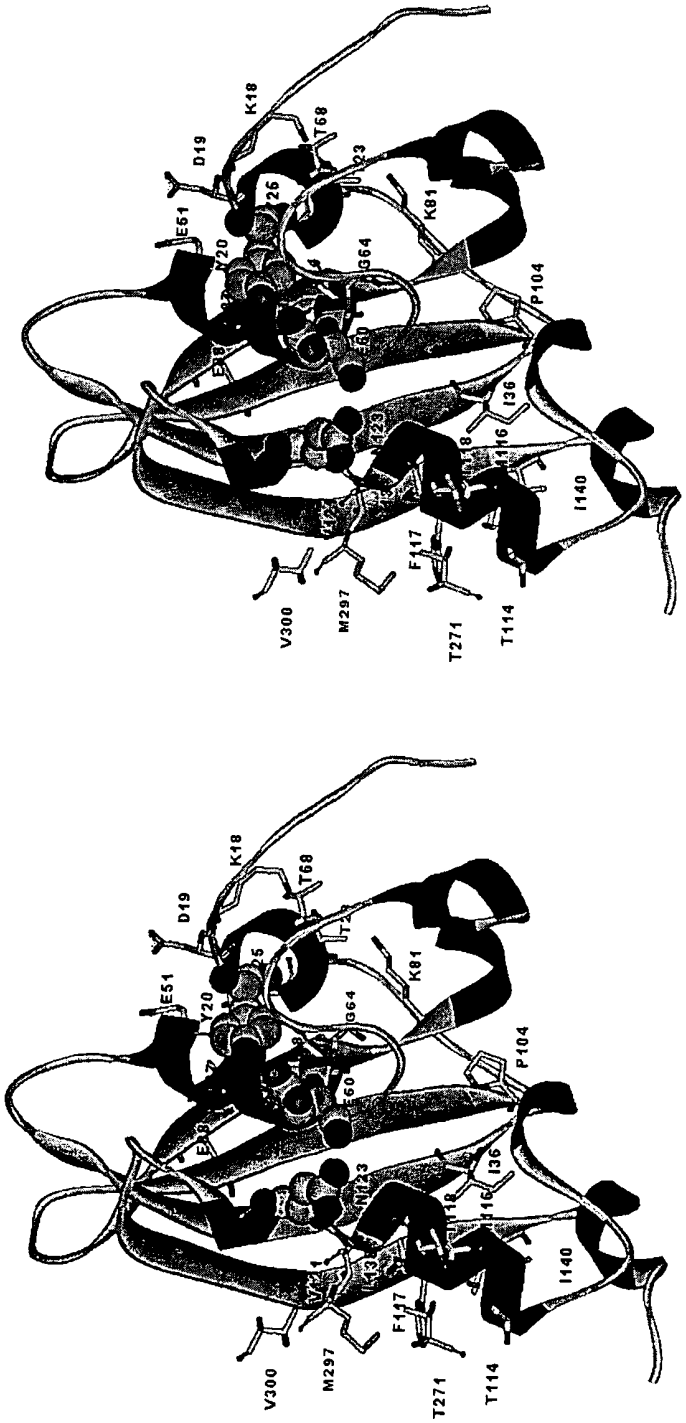

K18
D19
T68
E51
Y20
K81
G64
P104
I36
I116
I140
F117
T114
T271
M297
V300

FIG. 13

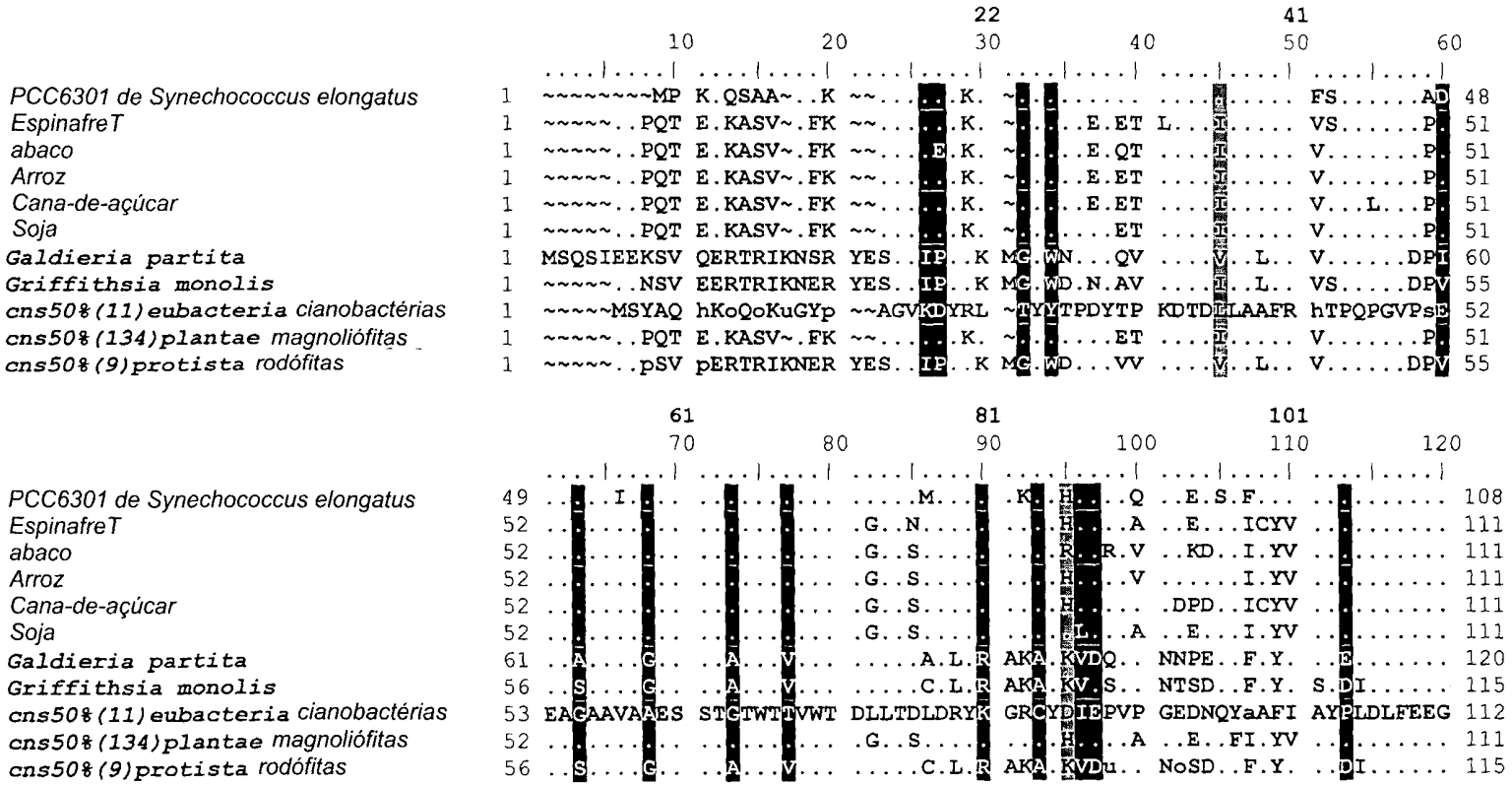

PCC6301 de Synechococcus elongatus
EspinafreT
abaco
Arroz
Cana-de-açúcar
Soja
Galdieria partita
Griffithsia monolis
cns50%(11)eubacteria cianobactérias
cns50%(134)plantae magnoliófitas
cns50%(9)protista rodófitas

FIG. 14A

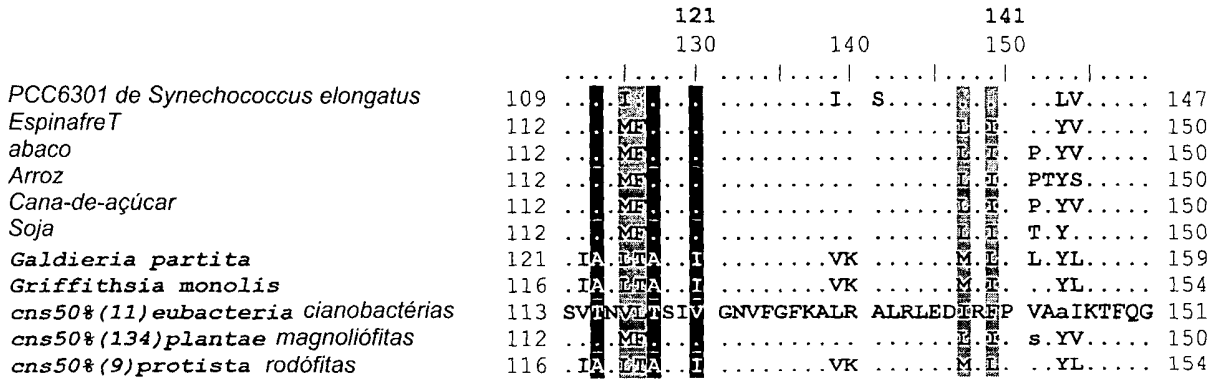

121
130
140
141
150
PCC6301 de Synechococcus elongatus 109 ..A.LTA..I ....I. S.....M.L. ..LV..... 147
EspinafreT 112 ..A.MF..I ....M.L. ..YV..... 150
abaco 112 ..A.MF..I ....M.L. P.YV..... 150
Arroz 112 ..A.MF..I ....M.L. PTYS..... 150
Cana-de-açúcar 112 ..A.MF..I ....M.L. P.YV..... 150
Soja 112 ..A.MF..I ....M.L. T.Y..... 150
Galdieria partita 121 .IA.LTA..I ....VK .....M.L. L.YL..... 159
Griffithsia monolis 116 .IA.LTA..I ....VK .....M.L. ..YL..... 154
cns50%(11)eubacteria cianobactérias 113 SVTNVLTSIV GNVFGFKALR ALRLEDMRFP VAaIKTFQG 151
cns50%(134)plantae magnoliófitas 112 ..A.MF..I ....M.L. s.YV..... 150
cns50%(9)protista rodófitas 116 .IA.LTA..I ....VK .....M.L. ..YL..... 154

FIG. 14B

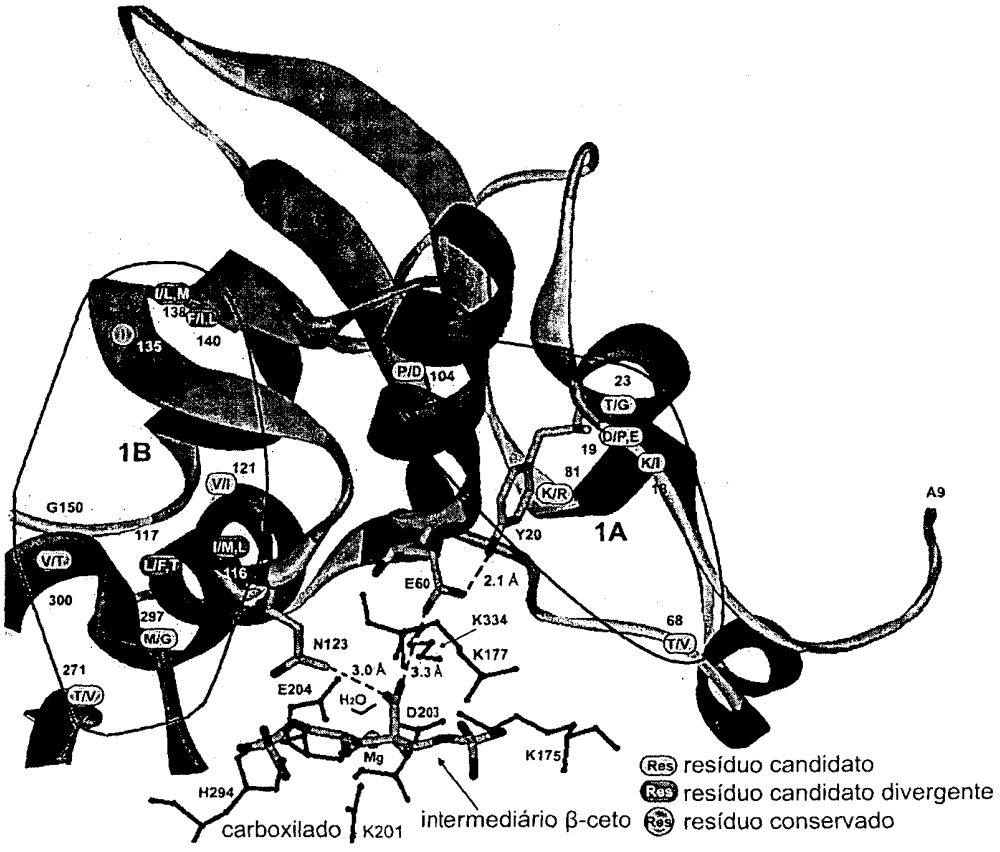

1B
1A
G150
117
121
135
140
104
23
19
81
16
A9
Y20
E60
2.1 Å
K334
N123
K177
3.0 Å
3.3 Å
E204
H2O
D203
Mg
K175
H294
carboxilado
K201
intermediário β-ceto
68
271
300
297
116
resíduo candidato
resíduo candidato divergente
resíduo conservado

FIG. 15

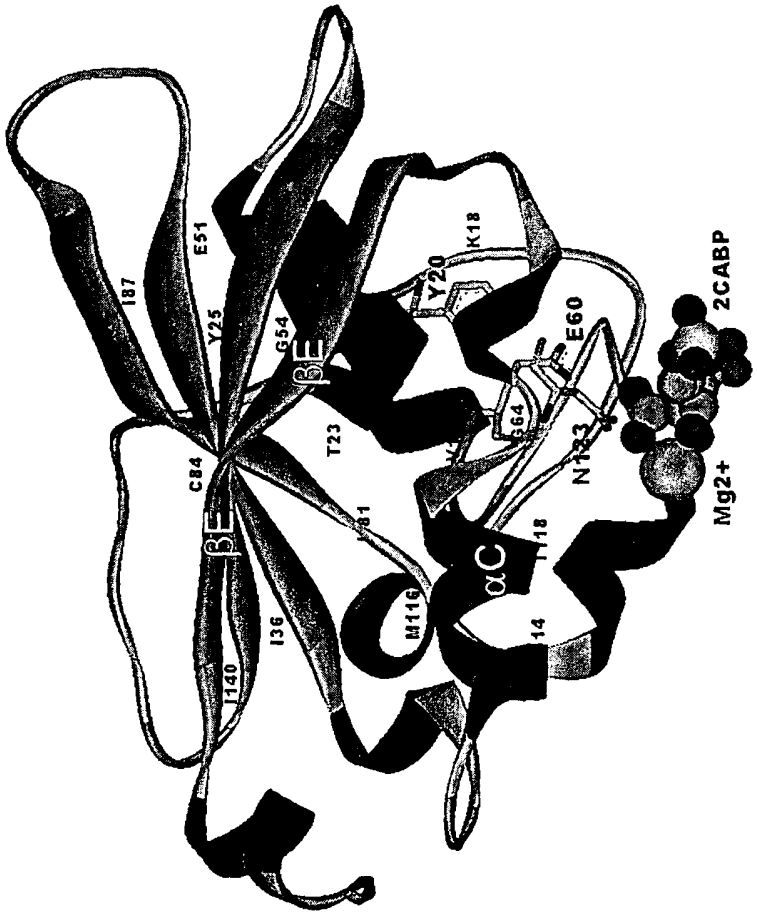

I87
E51
Y25
G54
βE
Y20
K18
E60
T23
C84
βF
αC
M116
I36
I140
Mg2+
2CABP

FIG. 16A

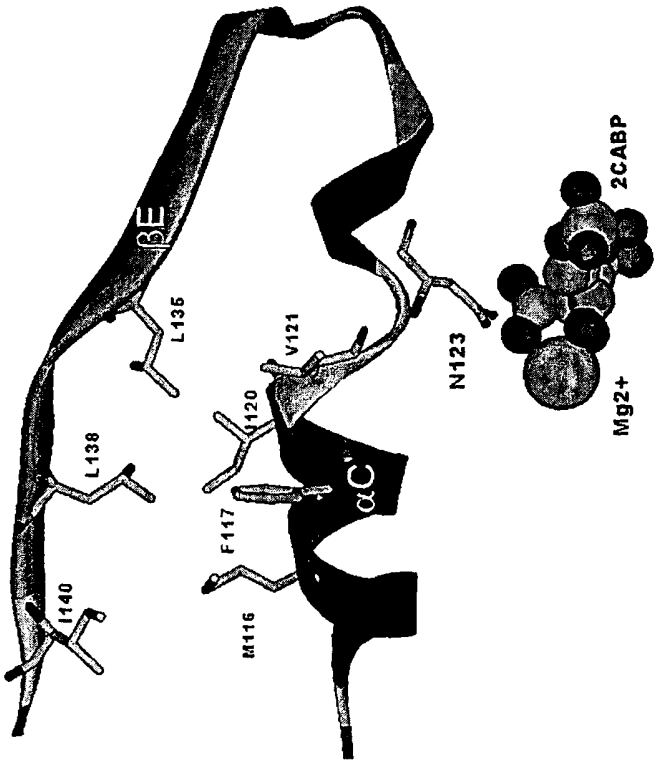


FIG. 16B

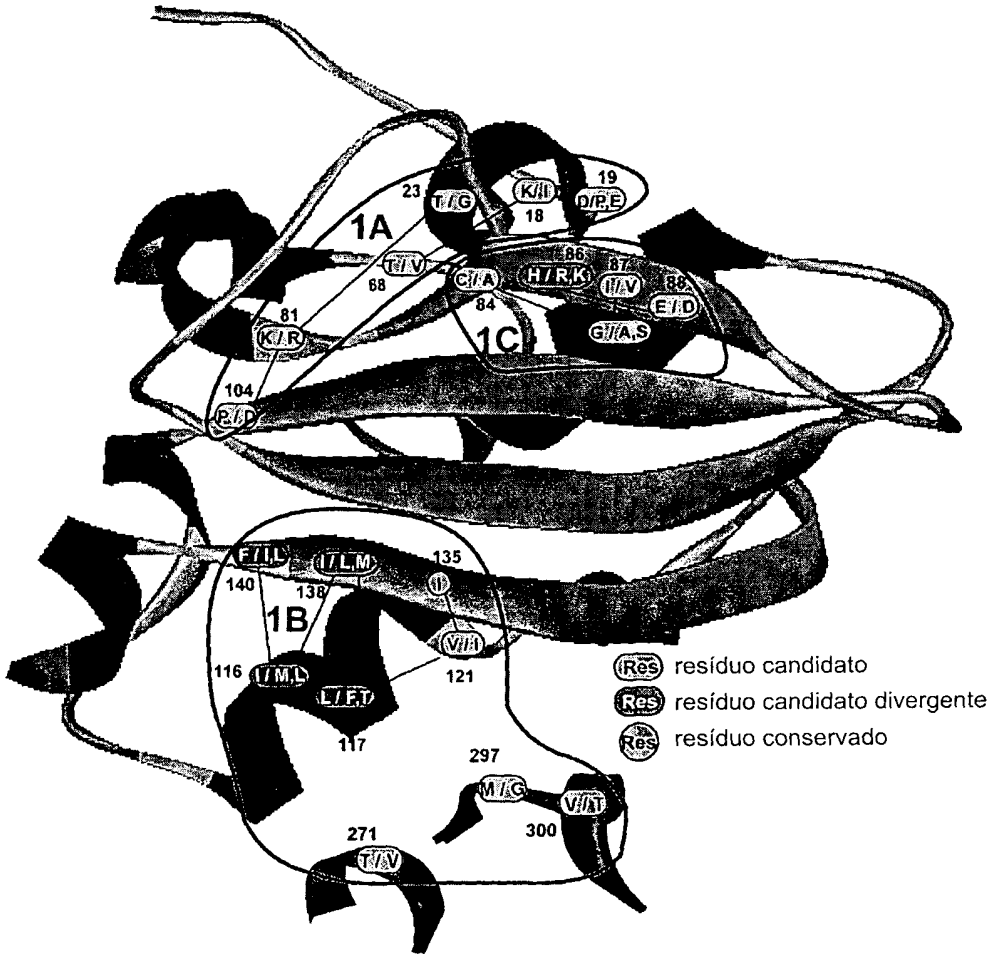

1A
1B
1C
23
T/G
K/I
19
D/P,E
18
T/V
68
C/A
84
H/R,K
86
87
I/V
E/D
81
K/R
G/A,S
104
P/D
F/I,L
I/L,M
135
140
138
V/I
116
I/M,L
L/F,T
121
117
297
M/G
V/T
271
T/V
300
Res resíduo candidato
Res resíduo candidato divergente
Res resíduo conservado

FIG. 17

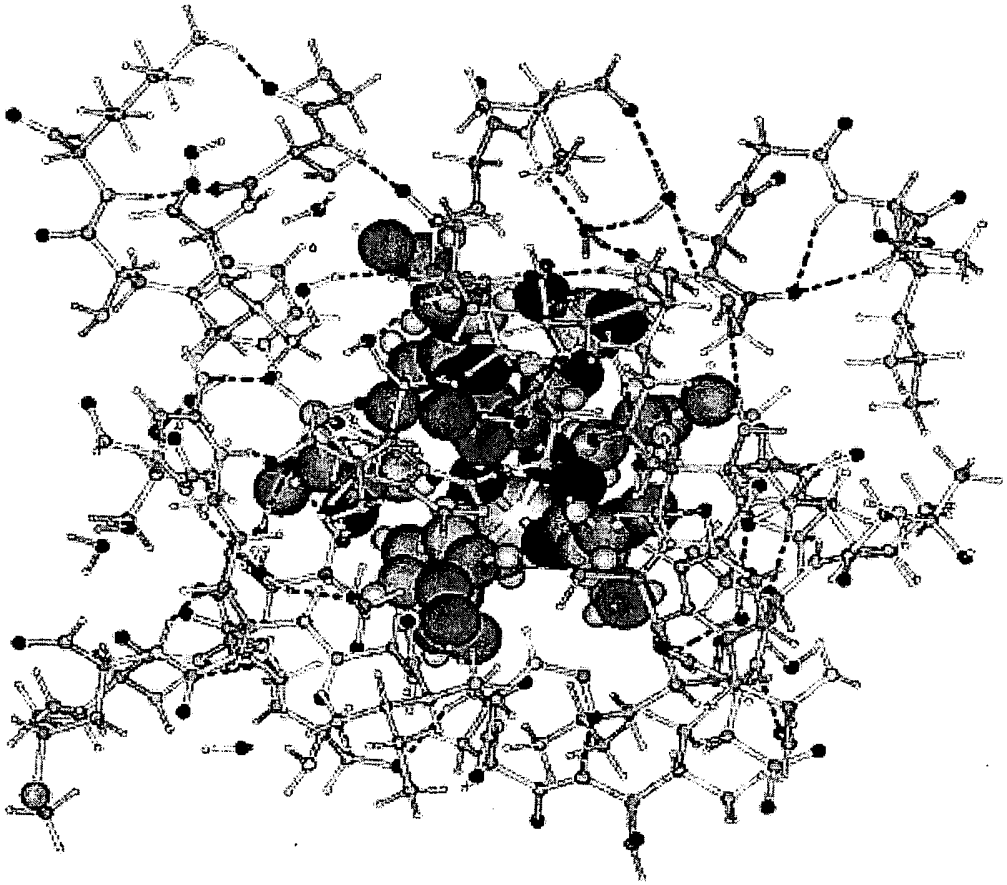


FIG. 18

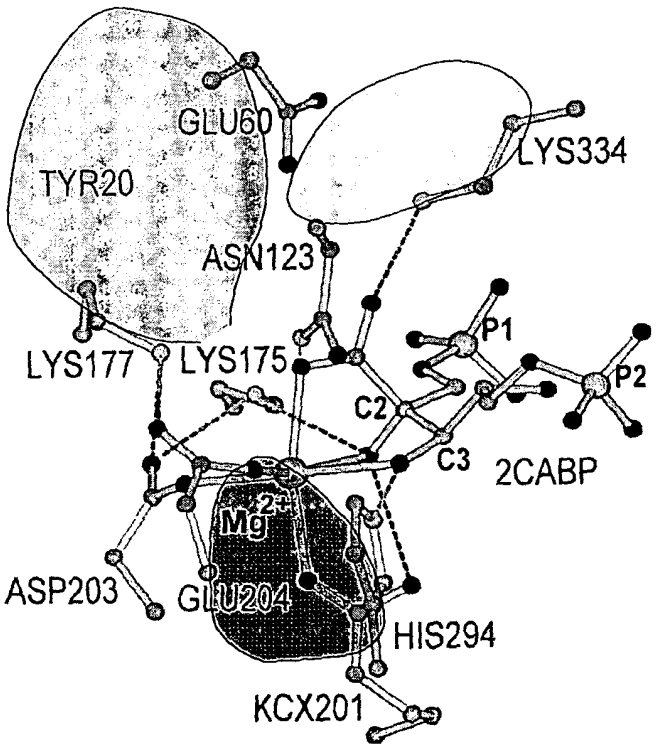

GLU60
LYS334
TYR20
ASN123
P1
LYS177
LYS175
P2
C2
C3
2CABP
Mg2+
ASP203
GLU204
HIS294
KCX201

FIG. 19

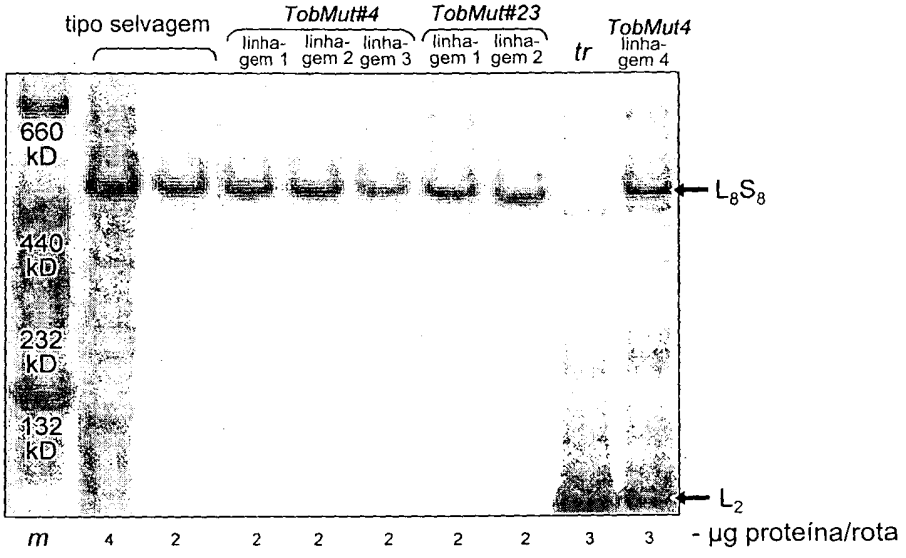

tipo selvagem
TobMut#4
TobMut#23
TobMut4
linha-gem 1
linha-gem 2
linha-gem 3
linha-gem 1
linha-gem 2
tr
linha-gem 4
660 kD
440 kD
232 kD
132 kD
L8S8
L2
m
4
2
2
2
2
2
2
2
3
3
- µg proteína/rota

FIG. 20