

(19) 日本国特許庁(JP)

(12) 公表特許公報(A)

(11) 特許出願公表番号

特表2005-505250

(P2005-505250A)

(43) 公表日 平成17年2月24日(2005.2.24)

(51) Int.Cl. ⁷	F I	テーマコード (参考)
C 1 2 N 15/09	C 1 2 N 15/00	Z N A A 2 B 0 3 0
A 0 1 H 5/00	A 0 1 H 5/00	A 2 B 1 5 0
A 2 3 K 1/16	A 2 3 K 1/16	3 O 4 C 4 B 0 1 8
C 0 7 K 19/00	C 0 7 K 19/00	4 B 0 2 4
C 1 2 N 9/02	C 1 2 N 9/02	Z T D 4 B 0 5 0
	審査請求 有 予備審査請求 未請求	(全 305 頁) 最終頁に続く

(21) 出願番号	特願2002-587384 (P2002-587384)	(71) 出願人	501154035
(86) (22) 出願日	平成14年5月6日 (2002.5.6)		ゼンコー
(85) 翻訳文提出日	平成15年10月31日 (2003.10.31)		X E N C O R
(86) 国際出願番号	PCT/US2002/014358		アメリカ合衆国91016カリフォルニア
(87) 国際公開番号	W02002/090300		州モンロピア、ウエスト・レモン・アベニ
(87) 国際公開日	平成14年11月14日 (2002.11.14)		ュー111番
(31) 優先権主張番号	60/289, 029	(71) 出願人	501016559
(32) 優先日	平成13年5月4日 (2001.5.4)		シンジェンタ・パティシペーションズ・ア
(33) 優先権主張国	米国 (US)		クチェングゼルシャフト
(31) 優先権主張番号	60/370, 609		S y n g e n t a P a r t i c i p a t
(32) 優先日	平成14年4月5日 (2002.4.5)		i o n s A G
(33) 優先権主張国	米国 (US)		スイス国, 4058 バーゼル, シュバル
(31) 優先権主張番号	60/376, 682		ツバルトアレー 215
(32) 優先日	平成14年4月29日 (2002.4.29)	(74) 代理人	100062144
(33) 優先権主張国	米国 (US)		弁理士 青山 稔
			最終頁に続く

(54) 【発明の名称】 核酸およびチオレドキシンレダクターゼ活性を有するタンパク質

(57) 【要約】

本発明は、変異体の少なくとも1つの物理的、化学的または生物学的特性が、野生型タンパク質と比較したときに、特異的および望まれる態様で変化している、機能性チオレドキシンレダクターゼ変異体を生成するための種々の方法の使用に関する。

【特許請求の範囲】

【請求項 1】

コンピューターの突然変異誘発を含むチオレドキシンレダクターゼのコファクター特異性を変化させるための方法。

【請求項 2】

チオレドキシンレダクターゼのコファクター特異性を変化させるための請求項1の方法であって、

(a)アミノ酸位置を含むチオレドキシンレダクターゼ(TR)スキャフォールドタンパク質の座標のセットを入力し、

(b)少なくとも1つのタンパク質設計サイクルを適用し、そして

(c)変化したコファクター依存性を有する候補変異体タンパク質のセットを生成させることを含む方法。

10

【請求項 3】

当該TRスキャフォールドタンパク質が、*E. coli*、*Bacillus subtilis*、*Mycobacterium leprae*、*Sarccharomyces*、*Neurospora crassa*、*Arabidopsos*、およびヒトからなる群より選択される、請求項2の方法。

【請求項 4】

当該変異体TRの当該コファクター特異性が、NADPHまたはNADHである、請求項1または2の方法。

【請求項 5】

当該変異体TRの当該コファクター特異性が、NADHに切り替えられている、請求項1または2の方法。

20

【請求項 6】

当該変異体TRの当該コファクター特異性が変化し、その結果当該変異体がNADHと比較してNADPHに優先的に結合する、請求項1または2の方法。

【請求項 7】

当該変異体TRの当該コファクター特異性が変化し、その結果当該変異体が、NADPHと比較して、NADHに優先的に結合する、請求項1または2の方法。

【請求項 8】

当該変異体TRの当該コファクター特異性が変化し、その結果当該変異体が、野生型TRタンパク質と比較して、NADPHについての改善された触媒効率を示す、請求項1または2の方法。

30

【請求項 9】

チオレドキシンレダクターゼの基質特異性を変化させるための方法であって、

(a)アミノ酸位置を含むチオレドキシンレダクターゼスキャフォールドタンパク質の座標のセットを入力し、

(b)少なくとも1つのタンパク質設計サイクルを適用し、そして

(c)変化した基質依存性を有する候補変異体タンパク質のセットを生成させることを含む方法。

【請求項 10】

当該変異体TRタンパク質が、*E. coli*、*Bacillus subtilis*、*Mycobacterium leprae*、*Sarccharomyces*、*Neurospora crassa*、*Arabidopsos*、およびヒトからなる群より選択される、生物から取得されるチオレドキシンタンパク質を還元する、変異体チオレドキシンレダクターゼ(TR)タンパク質。

40

【請求項 11】

当該変異体タンパク質が第2タンパク質に融合されている請求項1または2の変異体TRタンパク質であって、当該第2タンパク質が野生型TRタンパク質、チオレドキシン、または変異体TRタンパク質である、タンパク質。

【請求項 12】

当該変異体タンパク質がリンカーを介して当該第2タンパク質に融合されている請求項11

50

の変異体TRタンパク質。

【請求項 1 3】

当該変異体TRタンパク質が野生型ArabidopsisTRタンパク質と比較して、1ないし3つのアミノ酸置換を有する、請求項1又は2の変異体TRタンパク質。

【請求項 1 4】

当該アミノ酸置換が、位置A4、A5およびA6から選択される、請求項13の変異体TRタンパク質。

【請求項 1 5】

当該アミノ酸置換が、RA4W、RA5L、RA5M、RA5I、RA5F、RA5V、RA5Y、RA6T、RA6S、RA6Q、RA6G、およびRA6Nからなる置換の群より選択される、請求項14の変異体TRタンパク質。

10

【請求項 1 6】

アミノ酸置換RA4WおよびRA6Tを含む、請求項15の変異体TRタンパク質。

【請求項 1 7】

アミノ酸置換がRA4W、RA5L、およびRA6Nを含む、請求項15の変異体TRタンパク質。

【請求項 1 8】

アミノ酸置換RA5YおよびRA6Nを含む、請求項15の変異体TRタンパク質。

【請求項 1 9】

アミノ酸置換RA4W、RA5F、およびRA6Qを含む、請求項 1 5 の変異体タンパク質。

【請求項 2 0】

標的タンパク質のコファクター特異性を変化させるための方法であって、

20

(a) アミノ酸位置を含む、スキャフォールドタンパク質のための座標のセットを入力し、

(b) 少なくとも1つのタンパク質設計サイクルを適用しそして

(c) 変化したコファクター特異性を有する候補変異体タンパク質のセットを生成することを含む方法。

【請求項 2 1】

当該タンパク質設計サイクルがタンパク質設計オートメーション (PDA (登録商標)) を含む、請求項 1、2、9 または 20 の方法。

【請求項 2 2】

当該タンパク質設計サイクルが、配列予測アルゴリズムを含む、請求項 1、2、9、または 20 の方法。

30

【請求項 2 3】

当該タンパク質設計サイクルが、力場計算を含む、請求項 1、2、9、または 20 の方法。

【請求項 2 4】

式Iの単離ポリペプチド分子を含む変異体チオレドキシンレダクターゼ (TR) タンパク質；

(I) S₁-A₁-A₂-S₂-A₃-A₄-A₅-S₃-A₆-S₄

[式中

a) S₁は配列番号1、配列番号2、配列番号3、配列番号4、配列番号5、配列番号6、および配列番号7、またはそれに実質的類似性を有する配列からなる群より選択されるポリペプチド配列を含む；

40

b) S₂は配列番号8、配列番号9、配列番号10、配列番号11、配列番号12、配列番号13、および配列番号14、またはそれに実質的類似性を有する配列からなる群より選択されるポリペプチド配列を含む；

c) S₃は配列番号15、配列番号16、配列番号17、配列番号18、配列番号19、配列番号20、および配列番号21、またはそれに実質的類似性を有する配列からなる群より選択されるポリペプチド配列を含む；

d) S₄は配列番号22、配列番号23、配列番号24、配列番号25、配列番号26、配列番号27、および配列番号28、またはそれに実質的類似性を有する配列からなる群より選択されるポリペプチド配列を含む；

e) A₁はセリン、バリン、グリシン、アラニン、ロイシン、イソロイシン、メチオニン、フ

50

フェニルアラニン、およびトリプトファンからなる群より選択されるアミノ酸部分である；
f) A_2 はアラニン、グリシン、バリン、ロイシン、イソロイシン、メチオニン、フェニルアラニン、およびトリプトファンからなる群より選択されるアミノ酸部分である；

g) A_3 はヒスチジン、アスパラギン酸、グルタミン酸、アルギニン、ロイシン、セリン、トレオニン、システイン、アスパラギン、グルタミン、およびチロシンからなる群より選択されるアミノ酸部分である；

h) A_4 はアルギニン、アラニン、グリシン、バリン、ロイシン、イソロイシン、メチオニン、フェニルアラニン、およびトリプトファンからなる群より選択されるアミノ酸部分である；

i) A_5 はアルギニン、アスパラギン、グルタミン、アスパラギン酸、グルタミン酸、システイン、セリン、トレオニン、およびリジンからなる群より選択されるアミノ酸部分である； 10

j) A_6 はアルギニン、グルタミン酸、アスパラギン、グルタミン、アスパラギン酸、システイン、セリン、トレオニン、およびリジンからなる群より選択されるアミノ酸部分である；

ただし少なくとも

A_1 はセリンでない；

A_2 はアラニンでない；

A_3 はヒスチジンでない；

A_4 はアルギニンでない； 20

A_5 はアルギニンでない；または

A_6 はアルギニンでない。]

【請求項 2 5】

請求項 24 のポリペプチド分子；ここで S_1 は配列番号 1、配列番号 2、配列番号 3、配列番号 4、配列番号 5、配列番号 6、および配列番号 7 からなる群より選択される配列を有するポリペプチド配列からなる。

【請求項 2 6】

請求項 24 のポリペプチド分子；ここで S_2 は配列番号 8、配列番号 9、配列番号 10、配列番号 11、配列番号 12、配列番号 13、および配列番号 14 からなる群より選択される配列を有するポリペプチド配列からなる。 30

【請求項 2 7】

請求項 24 のポリペプチド分子；ここで S_3 は配列番号 15、配列番号 16、配列番号 17、配列番号 18、配列番号 19、配列番号 20、および配列番号 21 からなる群より選択される配列を有するポリペプチド配列からなる。

【請求項 2 8】

請求項 24 のポリペプチド分子；ここで S_4 は配列番号 22、配列番号 23、配列番号 24、配列番号 25、配列番号 26、配列番号 27、および配列番号 28 からなる群より選択される配列を有するポリペプチド配列からなる。

【請求項 2 9】

請求項 24 のポリペプチド分子；ここで S_1 は配列番号 1 に示すポリペプチド配列であり、 S_2 は配列番号 8 に示すポリペプチド配列であり、 S_3 は配列番号 15 に示すポリペプチド配列であり、かつ S_4 は配列番号 22 に示すポリペプチド配列である。 40

【請求項 3 0】

請求項 24 のポリペプチド分子；ここで S_1 は配列番号 2 に示すポリペプチド配列であり、 S_2 は配列番号 9 に示すポリペプチド配列であり、 S_3 は配列番号 16 に示すポリペプチド配列であり、かつ S_4 は配列番号 23 に示すポリペプチド配列である。

【請求項 3 1】

請求項 24 のポリペプチド分子；ここで S_1 は配列番号 3 に示すポリペプチド配列であり、 S_2 は配列番号 10 に示すポリペプチド配列であり、 S_3 は配列番号 17 に示すポリペプチド配列であり、かつ S_4 は配列番号 24 に示すポリペプチド配列である。 50

【請求項 3 2】

請求項24のポリペプチド分子；ここで S_1 は配列番号4に示すポリペプチド配列であり、 S_2 は配列番号11に示すポリペプチド配列であり、 S_3 は配列番号18に示すポリペプチド配列であり、かつ S_4 は配列番号25に示すポリペプチド配列である。

【請求項 3 3】

請求項24のポリペプチド分子；ここで S_1 は配列番号5に示すポリペプチド配列であり、 S_2 は配列番号12に示すポリペプチド配列であり、 S_3 は配列番号19に示すポリペプチド配列であり、かつ S_4 は配列番号26に示すポリペプチド配列である。

【請求項 3 4】

請求項24のポリペプチド分子；ここで S_1 は配列番号6に示すポリペプチド配列であり、 S_2 は配列番号13に示すポリペプチド配列であり、 S_3 は配列番号20に示すポリペプチド配列であり、かつ S_4 は配列番号27に示すポリペプチド配列である。 10

【請求項 3 5】

請求項24のポリペプチド分子；ここで S_1 は配列番号7に示すポリペプチド配列であり、 S_2 は配列番号14に示すポリペプチド配列であり、 S_3 は配列番号21に示すポリペプチド配列であり、かつ S_4 は配列番号28に示すポリペプチド配列である。

【請求項 3 6】

請求項24のポリペプチド分子；ここで A_1 はバリン、アラニン、およびロイシンからなる群より選択されるアミノ酸部分である。

【請求項 3 7】

請求項24のポリペプチド分子；ここで A_2 はグリシン、バリン、およびロイシンからなる群より選択されるアミノ酸部分である。 20

【請求項 3 8】

請求項24のポリペプチド分子；ここで A_3 はアスパラギン酸、グルタミン酸、アスパラギン、およびグルタミンからなる群より選択されるアミノ酸部分である。

【請求項 3 9】

請求項24のポリペプチド分子；ここで A_4 はアラニン、グリシン、バリン、ロイシン、イソロイシン、およびメチオニンからなる群より選択されるアミノ酸部分である。

【請求項 4 0】

請求項24のポリペプチド分子；ここで A_5 はアスパラギン、グルタミン、アスパラギン酸、およびグルタミン酸からなる群より選択されるアミノ酸部分である。 30

【請求項 4 1】

請求項24のポリペプチド分子；ここで A_6 はグルタミン酸、グルタミン、アスパラギン酸、およびアスパラギンからなる群より選択されるアミノ酸部分である。

【請求項 4 2】

請求項24のポリペプチド分子；ここで A_1 はバリンである。

【請求項 4 3】

請求項24のポリペプチド分子；ここで A_2 はグリシンである。

【請求項 4 4】

請求項24のポリペプチド分子；ここで A_3 はアスパラギン酸である。 40

【請求項 4 5】

請求項24のポリペプチド分子；ここで A_4 はアラニンである。

【請求項 4 6】

請求項24のポリペプチド分子；ここで A_5 はアスパラギンである。

【請求項 4 7】

請求項24のポリペプチド分子；ここで A_6 はグルタミン酸である。

【請求項 4 8】

請求項のポリペプチド分子24；ここで当該分子が*E.coli*、*Bacillus subtilis*、*Mycobacterium leprae*、*Sarccharomyces*、*Neurospora crassa*、*Arabidopsis*、およびヒトからなる群より選択される生物から取得されるチオレドキシントタンパク質を還元する、請求項24の 50

ポリペプチド分子。

【請求項 49】

請求項24のポリペプチド分子；ここで当該チオレドキシンの還元がコファクターの存在下で起こる。

【請求項 50】

請求項24のポリペプチド分子；ここで当該コファクターがNADPHまたはNADHである。

【請求項 51】

請求項24のポリペプチド分子；ここで当該コファクターはNADHである。

【請求項 52】

請求項24のポリペプチド分子；ここで当該ポリペプチドがNADHについてよりも、NADPHについて100倍大きいアフィニティを示す。 10

【請求項 53】

請求項24のポリペプチド分子；ここで当該ポリペプチドがNADHについてよりも、NADPHについて50倍大きいアフィニティを示す。

【請求項 54】

請求項24のポリペプチド分子；ここで当該ポリペプチドがNADHについてよりも、NADPHについて25倍大きいアフィニティを示す。

【請求項 55】

請求項24の単離ポリペプチド分子；ここで当該ポリペプチドが、第2ポリペプチドに融合されている。 20

【請求項 56】

請求項55のポリペプチド分子；ここで当該ポリペプチド当該ポリペプチドにリンカーを介して融合されている。

【請求項 57】

請求項56のポリペプチド分子；ここで当該リンカーが約5と約50個の間のアミノ酸を有するポリペプチド配列を含む。

【請求項 58】

請求項56のポリペプチド分子；ここで当該リンカーが約10と約40個の間のアミノ酸を有するポリペプチド配列を含む。

【請求項 59】

請求項56のポリペプチド分子；ここで当該リンカーが約15と約25個の間のアミノ酸を有するポリペプチド配列を含む。 30

【請求項 60】

請求項56のポリペプチド分子；ここで当該第2ポリペプチドがチオレドキシンである。

【請求項 61】

請求項56のポリペプチド分子；ここで当該ポリペプチドが、第3ポリペプチドにさらに融合されている。

【請求項 62】

請求項56のポリペプチド分子；ここで当該ポリペプチドが第3ポリペプチドにリンカーを介して融合されている。 40

【請求項 63】

請求項56または62のポリペプチド分子；ここで当該リンカーは、約5と約100kDaの間の分子量を有するポリペプチド配列を含む。

【請求項 64】

請求項56または62のポリペプチド分子；ここで当該リンカーが約25と約70kDaの間の分子量を有するポリペプチド配列を含む。

【請求項 65】

請求項56または62のポリペプチド分子；ここで当該リンカーが約25と約45kDaの間の分子量を有するポリペプチド配列を含む。

【請求項 66】

請求項56または62のポリペプチド分子；ここで当該ポリペプチドがオレオシンである。

【請求項67】

修飾されたTRタンパク質を有する植物を生産する方法であって、

(a)アミノ末端クロロプラスト移行ペプチドを含む、請求項1または22の修飾されたチオレドキシシンレダクターゼ(TR)をコードするDNA分子に作動可能に連結された植物細胞中で機能性であるプロモーターを含む発現カセットを、植物細胞内に導入し、形質転換植物細胞を得ること、および

(b)当該形質転換植物細胞を再生し、区別される形質転換植物を提供すること、ここで当該植物での修飾されたTRタンパク質をコードするDNA分子の発現が、非形質転換植物と比較して、コファクター特異性が変化していること

10

を含む方法。

【請求項68】

請求項67の方法であって、ここで当該形質転換植物が、当該コファクター特異性がNADPHまたはNADHである修飾されたTRタンパク質を発現する、方法。

【請求項69】

請求項67の方法であって、ここで当該形質転換植物が、当該コファクター特異性がNADHに切り替えられている修飾されたTRタンパク質を発現する、方法。

【請求項70】

当該形質転換植物が、修飾されたTRタンパク質を発現し、ここで当該コファクター特異性が変化して、その結果当該修飾されたTRタンパク質が優先的に、NADHと比較してNADPHに結合する、請求項67の方法。

20

【請求項71】

当該形質転換植物が、修飾されたTRタンパク質を発現し、ここで当該コファクター特異性が変化して、その結果当該修飾されたTRタンパク質が、非形質転換植物における野生型TRタンパク質と比較して、NADPHについての改善された触媒効率を示す、請求項1または2の方法。

【請求項72】

請求項67の方法によって調製される形質転換植物。

【請求項73】

請求項72の当該形質転換植物の形質転換された種子。

30

【請求項74】

修飾されたチオレドキシシンレダクターゼ(TR)タンパク質を含む油体を作成するための方法であって、

a)請求項1または2の修飾されたTRタンパク質を細胞中で生産し、

b)修飾されたTRタンパク質および当該油体を連関できる油体ターゲティングタンパク質によって当該TRタンパク質と油体を連関させ、および

c)当該修飾されたTRタンパク質と連関された当該油体を取得することを含む方法。

【請求項75】

さらに

40

a)該油体を洗浄し、当該修飾されたTRタンパク質を含む、洗浄された油体を取得する、および

b)当該洗浄された油体を、乳化物中に製剤化することを含む、請求項74の方法。

【請求項76】

当該油体を、非アレルギー性食品の調製で使用する、請求項74の方法。

【請求項77】

当該油体を動物飼料の調製で使用する、当該飼料の消化可能性を改良する、請求項74の方法。

【発明の詳細な説明】

50

【0001】

この出願は、2001年3月4日出願の米国出願番号60/289029、2002年4月5日出願の米国出願番号60/370609、および2002年4月29日出願で出願番号は付与されていないDesjariaisおよびMuchhalによる仮出願、名称「新規核酸およびチオレドキシンレダクターゼ活性を有するタンパク質」の出願日の利益を請求する。

【0002】

発明の分野

本発明は、変異体の少なくとも1つの物理的、化学的または生物学的特性が、野生型タンパク質と比較したときに、特異的および望ましい態様で変化された機能性チオレドキシンレダクターゼ変異体を生成するための種々の方法の使用に関する。

10

【0003】

発明の背景

チオレドキシン、小ジチオールタンパク質は、動物および植物源からの、主要な食品タンパク質、アレルギー性タンパク質および特に広く使用される食品中に存在するアレルギー性タンパク質のための特異的還元剤である。ジスルフィド(S-S)結合を有するほとんどのタンパク質は、チオレドキシンによってスルフヒドリル(SH)レベルへ還元される。これらのタンパク質は、アレルギー的に活性であり、そして酸化(S-S)状態でより消化可能でない。還元されたとき(SH状態)、これらはそのアレルギー性を喪失し、および/またはより消化可能になる。多くの種子および禾穀類中に見出されるアルブミン、グロブリン、グリアジン、チオニンおよびグルテニンおよびまたミルク中に見出されるある数のタンパク質のようなタンパク質中のジスルフィド結合のチオレドキシン還元は重要である。例えばKiss, F. et al. (1991), Arch. Biochem. Biophys. 287:337-340; Johnson, T. C. et al. (1987), Plant Physiol. 85:446-451; Kasarda, D. D. et al. (1976), Adv. Cer. Sci. Tech. 1:158-236; およびOsborne, T. B. et al. (1993), Amer. Chem. J. 15:392-471; Shewry, P. R. et al. (1985), Adv. Cer. Sci. Tech. 7:1-83; Dahle, L. K. et al. (1966), Cereal Chem. 43:682-688; Garcia-Olmedo, F. et al. (1987), Oxford Surveys of Plant MolecularおよびCell Biology 4:275-335; Birk, Y. (1976), Meth. Enzymol. 45:695-739, およびLaskowski, M., Jr. et al. (1980), Ann. Reo. Biochem. 49:593-626; Weselake, R. J. et al. (1983), Plant Physiol. 72:809-812; Birk, Y. (1985), Int. J. Peptide タンパク質 Res. 25:113-131, およびBirk, Y. (1976), Meth. Enzymol. 45:695-739; Birk, Y. (1985), Int. J. Peptide タンパク質 Res. 25:113-131参照。

20

30

【0004】

加えて、チオレドキシンは、多くの毒性タンパク質例えば、ヘビ(Yang, C. C. (1967) Biochim. Biophys. Acta. 133:346-355; Howard, B. D. et al. (1977) Biochemistry 16:122-125)、ミツバチ、サソリ(Watt, D. D. et al. (1972) Toxicon 10:173-181)、細菌神経毒素破傷風およびボツリヌス症(Schiavo, G. et al. (1990) InfectionおよびImmunity 58:4136-4141; Kistner, A. et al. (1992) Naunyn-Schmiedeberg's Arch Pharmacol 345:227-234)中のジスルフィド結合を還元し、それによってそれらの毒性をいっしょに減少させ、またはある場合には排除する。

40

【0005】

チオレドキシンは、NADP-チオレドキシンレダクターゼ(生理学的条件)を介してニコチンアミドデニンジヌクレオチドリン酸(NADPH)によって、またはジチオスレイトール、化学的還元剤によって活性化され(還元され)るとき、この還元を達成する。例えば引用によりここに全体を含める米国特許番号5952034参照。感作化イヌで実施された皮膚試験および喫食実験は、接種に先立ち還元されたチオレドキシンでの食品の処理は、食品のアレルギー性を排除または減少することを示した。研究はまた、チオレドキシンによる還元が続いて、ペプシンおよびトリプシンによる食品および食品たんぱく質の増加消化を示した。

【0006】

50

こうして、チオレドキシンレダクターゼを使用する、特性タンパク質の毒性を減少させる、食品のアレルゲン性を減少させ得る、そして食品の消化可能性を増加する、効率的、低コスト方法を開発することが望まれる。

【0007】

発明の要約

前概説目的にしたがって、本発明は、チオレドキシンレダクターゼのコファクター特異性を変化させる方法であって、アミノ酸位置を含む、チオレドキシンレダクターゼスキヤフォールドタンパク質のための座標のセットを入力すること；少なくとも1つのタンパク質設計サイクルを適用すること、および変化したコファクター従属性を有する候補タンパク質のセットを生成することを含む方法を提供する。好ましくは、該スキヤフォールドタンパク質は、*E. coli*、*Bacillus subtilis*、*Mycobacterium Leprae*、*Sarccharomyces*、*Neurospora crassa*、アラビドプシス、およびヒトからなる生物の群より選択される。

【0008】

追加的側面では、変異体TRタンパク質のコファクター特異性は、NADPHまたはNADHである。好ましくは、コファクター特異性は、NADHに切り替えられる。加えて、NADHと比較してNADPHに優先的に結合する、NADPHと比較してNADNに優先的に結合する、両方のコファクターに同等に結合する他のTR変異体を生成する。他の実施態様では、1または他のコファクターまたは両方についての触媒効率性を変化させる。

【0009】

追加的側面では、変異体TRは、RA4W、RA5L、RA5M、RA5I、RA5F、RA5V、RA5Y、RA5A、RA5S、RA5C、RA5T、RA6T、RA6S、RA6Q、RA6G、およびRA6N、RA6D、RA6M、およびRA6Eからなる置換の群より選択されるアミノ酸置換を有する。

【0010】

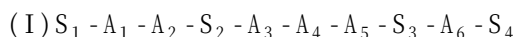
追加的側面では、本発明は、TRタンパク質の基質特異性を変化させるための方法であって、アミノ酸位置を含む、チオレドキシンレダクターゼスキヤフォールドタンパク質についての座標のセットを入力すること；少なくとも1つのタンパク質設計サイクルを適用すること、および変化した基質特異性を有する候補変異タンパク質ノセットを生成することを含む方法を提供する。

【0011】

追加的側面では、本発明は、標的タンパク質のコファクター特異性を変化させるための方法であって、アミノ酸位置を含む、チオレドキシンレダクターゼスキヤフォールドタンパク質についての座標のセットを入力すること；少なくとも1つのタンパク質設計サイクルを適用すること、および変化したコファクター特異性を有する候補変異タンパク質のセットを生成することを含む方法を提供する。

【0012】

追加的実施態様では、本発明は、変異体チオレドキシンレダクターゼ（TR）タンパク質であって、式I



式中

a) S_1 は配列番号1、配列番号2、配列番号3、配列番号4、配列番号5、配列番号6、および配列番号7からなる群より選択されるポリペプチド配列またはそれに実質的類似性を有する配列を含む、；

【0013】

b) S_2 は配列番号8、配列番号9、配列番号10、配列番号11、配列番号12、配列番号13、および配列番号14からなる群より選択されるポリペプチド配列またはそれに実質的類似性を有する配列を含む、；

【0014】

c) S_3 は配列番号15、配列番号16、配列番号17、配列番号18、配列番号19、配列番号20、および配列番号21からなる群より選択されるポリペプチド配列、またはそれに実質的類似性

10

20

30

40

50

を有する配列を含む；

【0015】

d)S₄は配列番号22、配列番号23、配列番号24、配列番号25、配列番号26、配列番号27、および配列番号28、からなる群より選択されるポリペプチド配列またはそれに実質的類似性を有する配列を含む；

【0016】

e)A₁はセリン、バリン、グリシン、アラニン、ロイシン、イソロイシン、メチオニン、フェニルアラニン、およびトリプトファンからなる群より選択されるアミノ酸部分であり、

【0017】

f)A₂はアラニン、グリシン、バリン、ロイシン、イソロイシン、メチオニン、フェニルアラニン、およびトリプトファンからなる群より選択されるアミノ酸部分であり； 10

【0018】

g)A₃はヒスチジン、アスパラギン酸、グルタミン酸、アルギニン、ロイシン、セリン、トレオニン、システイン、アスパラギン、グルタミン、およびチロシンからなる群より選択されるアミノ酸部分であり；

【0019】

h)A₄はアルギニン、アラニン、グリシン、バリン、ロイシン、イソロイシン、メチオニン、フェニルアラニン、およびトリプトファンからなる群より選択されるアミノ酸部分であり；

【0020】

i)A₅はアルギニン、アスパラギン、グルタミン、アスパラギン酸、グルタミン酸、システイン、セリン、トレオニン、およびリジンからなる群より選択されるアミノ酸部分であり； 20

【0021】

j)A₆はアルギニン、グルタミン酸、アスパラギン、グルタミン、アスパラギン酸、システイン、セリン、トレオニン、およびリジンからなる群より選択されるアミノ酸部分である；

【0022】

ただし少なくとも

A₁はセリンではない； 30

A₂はアラニンではない；

A₃はヒスチジンではない；

A₄はアルギニンではない；

A₅はアルギニンではない；または

A₆はアルギニンではない。

【0023】

追加的側面では、本発明は、植物細胞の油含量を変化させる方法であって、アミノ末端クロロプラスト移行ペプチドを含む請求項1または22の修飾されたチオレドキシシンレダクターゼ（TR）タンパク質をコードするDNA分子に作動可能に連結された植物細胞中で機能性のプロモーターを含む発現カセットを、植物細胞内に導入し、形質転換された植物細胞を得ること；および当該形質転換植物細胞を再生し、区別される形質転換植物を提供することを含む方法を提供し、ここで修飾されたTRタンパク質をコードするDNA分子の、当該植物での発現が、非形質転換植物と比較してコファクター特異性が変化している。 40

【図面の簡単な説明】

【0024】

【図1】図1はチオレドキシシンレダクターゼによって触媒される反応である。

【図2A】図2は高度に保存されるある数の種からのレダクターゼの活性部位ポケットを示す。もっとも普通のTR配列の幾つかである。最初のコラムは、GenbankID番号を列挙し、A1ないしA6は式I（以下に記載）に定義のアミノ酸を言及する。S2およびS3はA1ない 50

しA6を文意Rする配列ドメインであり、そしてまた式Iに定義される。

【図2B】図2は高度に保存されるある数の種からのレダクターゼの活性部位ポケットを示す。普通のグルタチオンレダクターゼ配列のいくつかである。

【図2C】図2は高度に保存されるある数の種からのレダクターゼの活性部位ポケットを示す。生物によってグループ化される定義された位置のそれぞれの天然配列多様性。

【図2D】図2は高度に保存されるある数の種からのレダクターゼの活性部位ポケットを示す。生物によってグループ化される定義された位置のそれぞれの天然配列多様性である。

【図2E】図2は高度に保存されるある数の種からのレダクターゼの活性部位ポケットを示す。既知のコファクター特異性および既知のアミノ酸配置である。

10

【図3-1】図3は式Iで使用し得る種々の配列である。

【図3-2】図3は式Iで使用し得る種々の配列である。

【図3-3】図3は式Iで使用し得る種々の配列である。

【図3-4】図3は式Iで使用し得る種々の配列である。

【図3-5】図3は式Iで使用し得る種々の配列である。

【図4】図4はハイスループットTRスクリーニング法の概略である。

【図5】図5はタンパク質精製戦略である。

【図6】図6はアラビドプシスNTR野生型レダクターゼと、NAD(P)Hの動力学である。

【図7】図7はNTR-1ライブラリー1から取得された変異体である。

【図8】図8はNTR-1ライブラリー2から取得された変異体である。

20

【図9A】図9AはNTR-1ライブラリー1およびNTR-2ライブラリー2から、設計された位置およびドック化されたコファクターである。

【図9B】図9BはNTR-1ライブラリー1およびNTR-2ライブラリー2から、設計された位置およびドック化されたコファクターである。

【図10】図10は4つのコンピューターのライブラリーからの変異体のスクリーニングから得たまとめである。

【図11A】図11Aは2つの変異体対野生型TRについての動力学パラメーターである。

【図11B】図11Bは2つの変異体対野生型TRについての動力学パラメーターである。

【図12】図12はNTR-1ライブラリー2設計から取得された最良変異体のまとめである。

30

【図13A】図13Aは高複雑性ランダムRRRライブラリーから取得した変異体の活性のまとめである。

【図13B】図13Bは高複雑性ランダムRRRライブラリーから取得した変異体の活性のまとめである。

【図13C】図13Cはこのライブラリーから取得した変異体のまとめである。

【図14】図14はクローンからの2つについてのコンピューターのモデルである。

【図15】図15は変異体のいくつかについての酵素的活性および動力学パラメータのまとめである。

【図16A-1】図16AはWVR変異体についての核酸配列である。

【図16B-1】図16BはWMG変異体についての核酸配列である。

40

【図16C-1】図16CはWIS変異体についての核酸配列である。

【図16D-1】図16DはWMS変異体についての核酸配列である。

【図16E-1】図16EはWLS変異体についての核酸配列である。

【図16F-1】図16FはWRT変異体についての核酸配列である。

【図16G-1】図16GはRYN変異体についての核酸配列である。

【図16H-1】図16HはRYN-A変異体についての核酸配列である。

【図16I-1】図16IはRFN変異体についての核酸配列である。

【図16J-1】図16JはRRR-WT核酸配列である。

【図16K-1】図16KはWVG変異体についての核酸配列である。

【図16L-1】図16LはWRS変異体についての核酸配列である。

50

【図 1 6 M-1】図 1 6 M は WFQ 変異体についての核酸配列である。

【図 1 6 N-1】図 1 6 N は NTR 野生型タンパク質についての核酸配列である。

【図 1 6 O-1】図 1 6 O は RYN-M 変異体についての核酸配列である。

【図 1 6 P-1】図 1 6 P は RYN-L 変異体についての核酸配列である。図 1 6 Q RYN-I 変異体についての核酸配列である。

【図 1 7 A】図 1 7 A は変異体のうちのいくつかとの、アラビドプシス NTR 野生型タンパク質のアラインメントである。

【図 1 7 B】図 1 7 B は変異体のうちのいくつかとの、アラビドプシス NTR 野生型タンパク質のアラインメントである。

【図 1 8】図 1 8 は実施例 1 にキシの重要な RRR ないし RNY 変化のコンピュータ的提示である。 10

【図 1 9】図 1 9 はタンパク質データベースから抜粋の NAD コンホメーションの小サンプルである。ボール-およびスティックモデルは、NAD TDF コンホーマーであり、これはその他のほとんどと異なるリブローズパックを有する。

【図 2 0】図 2 0 は PDA シミュレーションで利用されるライブラリー位置およびライブラリー 1 および 2 の生成である。

【図 2 1-1】図 2 1 はいくつかの野生型 TR タンパク質の配列アラインメントである。配列は、以下に対応する： 1) | P09625|TRXB_ECOLI; 2) | P80880|TRXB_BACSU; 3) | P46843|TRXB_MYCLE; 4) | P51978|TRXB_NEUCR; 5) | P29509|TRB1_YEAST; 6) | P38816|TRB2_YEAST; 7) | Q39243|TRB1_ARATH; 8) | Q39242|TRB2_ARATH; および、 9) | Q16881|TRXB_HUMAN 20

【図 2 1-2】図 2 1 はいくつかの野生型 TR タンパク質の配列アラインメントである。配列は、以下に対応する： 1) | P09625|TRXB_ECOLI; 2) | P80880|TRXB_BACSU; 3) | P46843|TRXB_MYCLE; 4) | P51978|TRXB_NEUCR; 5) | P29509|TRB1_YEAST; 6) | P38816|TRB2_YEAST; 7) | Q39243|TRB1_ARATH; 8) | Q39242|TRB2_ARATH; および、 9) | Q16881|TRXB_HUMAN

【図 2 1-3】図 2 1 はいくつかの野生型 TR タンパク質の配列アラインメントである。配列は、以下に対応する： 1) | P09625|TRXB_ECOLI; 2) | P80880|TRXB_BACSU; 3) | P46843|TRXB_MYCLE; 4) | P51978|TRXB_NEUCR; 5) | P29509|TRB1_YEAST; 6) | P38816|TRB2_YEAST; 7) | Q39243|TRB1_ARATH; 8) | Q39242|TRB2_ARATH; および、 9) | Q16881|TRXB_HUMAN 30

【図 2 2】図 2 2 はいくつかの野生型 TR タンパク質のアミノ酸配列である。配列は以下に対応する： A) | P09625|TRXB_ECOLI; B) | P80880|TRXB_BACSU; C) | P46843|TRXB_MYCLE; D) | P51978|TRXB_NEUCR; E) | P29509|TRB1_YEAST; F) | P38816|TRB2_YEAST; G) | Q39243|TRB1_ARATH; H) | Q39242|TRB2_ARATH; および、 I) | Q16881|TRXB_HUMAN。

【0025】

発明の詳細な記載

本発明は、変化したコファクター特異性を呈する変異体タンパク質および核酸の生成を指向する。該変異体タンパク質は、アルスの種々のアプローチを使用して生成し得る。例えば、慣行の突然変異誘発およびコンピュータ的加工アプローチである。コンピュータ的加工アプローチは、米国特許番号 6,188,965 および 6,296,312, 米国出願番号 09/419,351, 09/782,004, 09/927,79, および 09/877,695 であってすべて引用によりここに全体を含めるものに以前記載された。一般に、これらの出願は、種々のコンピュータ的モデル化システムを使用し、これは極度に安定なタンパク質の生成を可能とする。このやり方では、野生型タンパク質の変異体であって、野生型タンパク質と比較して変化したコファクター特異性を呈するものを生成する。 40

【0026】

本発明の方法を、その他を超えて 1 のコファクターについて優先性を呈する任意の酵素に適用し得る。例えば、酵素レダクターゼは、その他に対し 1 のコファクターについて優先性をしばしば呈する。加えて、本発明の方法を、標的タンパク質の基質特異性を変化させ 50

るために適用し得る。

【0027】

特に、本発明の方法を使用し、NADPHからNADHにコファクター優先性を変化させることができる。NADPHは高価な還元剤である。その費用は、食品アレルギーおよび毒液処置を減少させることにおけるチオレドキシンシステムの広い使用を禁止してきた。こうして、当分野で、より低いコストで、NADP-チオレドキシンレダクターゼ還元剤の使用として同じ結果を達成する他のシステムを見出す必要性がある。1つのそのようなシステムは、変化したコファクター特異性を有するチオレドキシンレダクターゼの変異体を生成する。

【0028】

本発明にしたがって、標的タンパク質のコファクター特異性を変化させる方法を提供する。ここに「変化」または文法的均等物によって、ポリペプチドの文脈では、ここで使用するように、天然に存在するタンパク質の対応する特性と比較して選択または検出できるポリペプチドの任意の特徴または属性をさらに言及する。これらのタンパク質は、限定しないが、コファクター特異性、細胞毒性活性、酸化的安定性、基質特異性、基質結合または触媒活性、温度安定性、アルカリ安定性、pH活性プロファイル、タンパク質溶解分解への抵抗性、動力学会合 (Kon) および解離 (Koff) 率、分泌のための能力、細胞の表面上のディスプレイ能力、オリゴマー化能力、シグナル化能力、細胞増殖を刺激する能力、細胞増殖を阻害する能力、アポトーシスを誘導する能力、ホスホリル化またはグリコシル化によって修飾される能力、疾患を処置する能力を含む。

10

【0029】

他に特定しない限り、前列挙特性のいずれかの実質的变化は、本発明の変異体ポリペプチドの特性を、標的タンパク質または野生型タンパク質の特性と比較するとき、好ましくは少なくとも20%、より好ましくは50%、より好ましくは少なくとも2倍増加または減少である。

20

【0030】

ここで「コファクター特異性」によって、酵素のコファクター優先性を変化させることを意味する。ここで「コファクター」によって、補酵素、例えば酸化／還元反応に参与するNADPH、NADHを意味する。こうして、標的タンパク質は、その他を超える1のコファクターについての優先性を呈する。本発明の方法を使用し、標的酵素のコファクター優先性を変化させ得、その結果より低い優先されたコファクターについての優先性は、20%、50%、100%、300%、500%、1000%、2000%まで増加する。例えばある数のレダクターゼ酵素は、NADHよりNADPHを優先する(WO 02/22526; WO 02.29019; Mittl, PR., et al., (1994) Protein Sci., 3: 1504-14; Banta, S., et al., (2002) Protein Eng., 15:131-140; すべて引用によりここに含めるのを参照)。NADPHの利用性は、インビボおよびインビトロの両方でしばしば限定的であるように、標的タンパク質の全体活性は、しばしば限定的である。コファクターとしてNADPHを好む標的タンパク質のために、より容易に利用可能であるコファクター、例えばNADHに標的タンパク質のコファクター特異性を変化させることが望ましい。

30

【0031】

好ましい実施態様では、標的タンパク質のコファクター特異性を切りかえる。「切りかえる」によってここに、標的タンパク質のコファクター優先性（例えば親和性）を、その他のコファクターに変化させる。好ましくは、1実施態様では、コファクター特異性を切り替えることによって、野生型酵素によって優先されるコファクターを有する活性を減少させ、一方より低い優先されたコファクターを有する活性を増加させる。例えば、もし標的タンパク質がNADPHを優先するならば、該優先性をNADHに切り替えることは、NADHを使用してネイティブNADPH依存性活性の少なくとも50%を有する変異体TRを得る。より好ましくは、変異体TRはNADHを使用して天然NADPH依存性活性の少なくとも75%を有する。より好ましくは変異体TRは、NADHを使用してネイティブNADPH活性の85%、95%、100%までを有する。代替的に、その他の実施態様では、代替的コファクター親和性は、優先されるコファクター親和性の現象を有することなく増加する。なおさらなる実施態様

40

50

では、両方のファクターについてのコファクター親和性は同時的に変化する。

【0032】

好ましくは、コファクターについての標的タンパク質の触媒効験は、増強される。ここに「触媒効験」によって、コファクターでの活性が顕著に改善されることを意味する。触媒効験は、いずれかの優先されるコファクターについて改善され、または、コファクター特異性が変化する場合の実施態様では、変化コファクターでの触媒効験は改善し得る。

【0033】

好ましい実施態様では、コファクターについての標的タンパク質の結合親和性は増強される。結合親和性の変化は、少なくとも5%またはより大きい増加または野生型標的タンパク質と比較したときの結合親和性の減少によって証拠付けられる。ある種の実施態様では、本発明の変異体タンパク質は、その他についてよりも1コファクターについての100倍より大きい親和性を示し得、一方他の実施態様では、変異体タンパク質は、その他についてよりも1コファクターについて50倍より大きい親和性を示す。または変異体タンパク質は、その他よりも1コファクターについて25倍より大きい親和性を示し得る。

10

【0034】

好ましい実施態様では、標的タンパク質の基質特異性は、変化する。例えば、標的タンパク質が典型的には、同じ種からの基質に作用するならば、標的タンパク質の基質特異性が変化し得、その結果変異体タンパク質は、他の種からの基質に作用し得る。

【0035】

よって、本発明は、標的タンパク質のコファクター特異性を変化させるための方法を指向する。ここに「標的タンパク質」または「スキャフォールドタンパク質」またはその文法的均等物は、タンパク質、ポリペプチド、オリゴペプチドおよびペプチドを含む少なくとも2つの共有結合アミノ酸を意味する。タンパク質は、天然存在アミノ酸およびペプチド結合、または全般に合成の方法に依存して、合成ペプチドミメチック構造すなわち「アナログ」例えばペプトイド[Simon et al., Proc. Natl. Acad. Sci. USA 89(20):9367-71(1992)参照]から作成し得る。こうしてここに使用するように「アミノ酸」または「ペプチド残基」は、天然存在および合成の両方を意味する。例えば、ホモフェニルアラニン、シトルリン、およびノレロイシンは本発明の目的のアミノ酸と考える。「アミノ酸」はまた、イミノ酸残基例えばプロリンおよびヒドロキシプロリンを含む。加えて、本発明の変異体タンパク質の構成要素を表すアミノ酸は、同じアミノ酸また反対キラリティによって置換し得る。こうしてL配置で天然に存在する任意のアミノ酸（これはまた、化学物質の構造に依存してRまたはSとして言及し得る）は、同じ化学的構造型のアミノ酸また反対キラリティであって一般にD-アミノ酸といわれ、しかし追加的にその組成および化学的配置に依存してRまたはSとして言及されるのと置換し得る。そのような誘導体は、大きく増加した安定性の特性を有し、そしてしたがって、より長いインビボ半減期を有し得る化合物の製剤で、経口、静脈内、筋肉内、腹腔内、局所、経腸、目内、またはその他の経路で投与したときに有利である。好ましい実施態様では、アミノ酸は(S)またはL配置である。もし非天然存在側鎖を使用するならば、非アミノ酸置換を使用し、例えばインビボ分解を防止または阻害し得る。非天然存在アミノ酸を含むタンパク質は合成し得、幾つかの場合は、組換え的に作成し得る。引用によりここに含める、van Hest et al., FEBS Lett 428:(1-2) 68-70 May 22 1998およびTang et al., Abstr. Pap Am. Chem. S218:U138-U138 Part 2 August 22, 1999参照。

20

30

40

【0036】

芳香族アミノ酸は、DまたはLナフチルアラニン、DまたはLフェニルグリシン、DまたはL-2-チエニルアラニン、DまたはL-1-、-2-、-3-または4-ピレニルアラニン、DまたはL-3-チエニルアラニン、DまたはL-(2-ピリジニル)-アラニン、DまたはL(3-ピリジニル)-アラニン、DまたはL-(2-ピラジニル)-アラニン、DまたはL(4-イソプロピル)-フェニルグリ親、D-(トリフルオロメチル)-フェニルグリシン、D-(トリフルオロメチル)-フェニルアラニン、D-p-フルオロフェニルアラニン、DまたはL-p-ビフェニルフェニルアラニン、DまたはL-p-メトキシビフェニルフェニルアラニン、DまたはL-2-インドール(アルキル)アラニン、

50

およびDまたはL-アルキルアラニンで置換し得る。ここで、アルキルは、置換または非置換メチル、エチル、プロピル、ヘキシル、ブチル、ペンチル、イソプロピル、イソブチル、sec-ブチル、sec-イソチル、イソ-ペンチル、非酸性アミノ酸であってC1-C20であるものであり得る。酸性アミノ酸は、非カルボキシレートアミノ酸であって、陰性荷電を維持するもの、およびその誘導体またはアナログ、例えば非限定的例は、(フォスフォノ)アラニン、グリシン、ロイシン、イソロイシン、トレオニン、またはセリン、または硫酸化(例えば-SO₃H置換3H)トレオニン、セリン、チロシンである。他の置換は、非天然ヒドロキシル化アミノ酸は、「アルキル」を任意の天然アミノ酸とあわせることにより作成し得る。用語「アルキル」はここで使用するように、分枝または非分枝飽和炭化水素基で1ないし24炭素原子のもの、例えばメチル、エチル、n-プロピル、イソプロピル、n-ブチル、イソブチル、t-ブチル、オクチル、デシル、テトラデシル、ヘキサデシル、エイコシル、テトラシル等である。

10

【0037】

アルキルは、窒素、酸素、および硫黄の原子を有するヘテロアルキルを含む。好ましいアルキル基は、ここで、1ないし12炭素原子を含む。塩基性アミノ酸は、天然存在アミノ酸リジン、アルギニン、オルニチン、シトルリンまたは(グアニジノ)-酢酸、または他の(グアニジノ)アルキル-酢酸の任意の位置でアルキル基で置換されい得る。ここで「アルキル」は、前定義の通りである。ニトリル誘導体(例えばCOOHの代わりにCN-部分を含む)は、またアスパラギンまたはグリタミンで置換し得、そしてメチオニンスルホキシドは、メチオニンで置換し得る。そのようなペプチド誘導体の製造の方法は、当業者に周知である。

20

【0038】

加えて、任意の変異体ペプチドの任意のアミド結合は、ケトメチレン部分によって置換できる。そのような誘導体は、酵素による分解に対する増加安定性を有することが予測される。したがって、経口、静脈内、筋肉内、腹腔内、局所、経腸、眼内または他の経路によって投与するように、増加インビボ半減期を有し得る化合物の製剤のために有利性を保有する。本発明の変異体ポリペプチドのアミノ酸の追加的アミノ酸修飾は、異化を含み得る：システイニル残基は、 α -ハロアセテート(および対応アミン)で、例えば2-クロロ酢酸またはクロロアセトアミドで置換し得、カルボキシメチルまたはカルボキシアミドメチル誘導体を与える。システイニル残基は、また化合物例えばプロモトリフルオロアセトン、 α -ブromo- β -(5-イミドゾイル)プロピオン酸、クロロアセチルフォスフェート、N-アルキルマレイミド、3-ニトロ-2-ピリジルジスルフィド、メチル-2-ピリジルジスルフィド、p-クロロメルクリベンゾエート、2-クロロメルクリ-4-ニトロフェノール、またはクロロ-7-ニトロベンゾ-2-オキサ-1,3-ジアゾールとの反応によって誘導体化し得る。ヒスチジル残基を、化合物、例えばジエチルプロカルボネートと、例えばpH5.5-7.0の反応によって誘導体化し得、というのは、この剤は、明かに、ヒスチジル側鎖に特異的であり、およびpara-ブromoフェナシルブロミドをまた使用し得る：例えばここで反応は、好ましくは0.1MナトリウムカゴジレートでpH6.0で実施する。リシニルおよびアミノ末端残基は、化合物、例えばコハク酸または他のカルボン酸無水物と反応し得る。これらの剤での誘導体化は、リシニル残基の変化を逆転する効果を有すると期待される。 α アミノ含有残基を誘導体化するための他の好適な試薬は、化合物例えばイミドエステル/例えばメチルピコリンイミデートとしてピリドキサールフォスフェート、ピリドキサール、ウロロポロヒドリド、トリニトロベンゼンスルホン酸；0-メチルイソウレア；2,4-ペンタンジオン；およびグリコシレートとのトランスアミナーゼ触媒性反応を含む。アルギニル残基は、1または幾つかの慣行の試薬、とりわけフェニルグリオキサール、2,3-ブタンジオン、1,2-シクロヘキサンジオン、およびニンヒドリンとの反応によって既知方法工程にしたがって修飾し得る。アルギニン残基の誘導体化は、反応がアルカリ条件で実施されることを要旨、というのは、グアニジン官能基のpKaが高いからである。さらに、これらの試薬は、リジンの基並びにアルギニンイプシロン-アミノ基と反応し得る。チロシル残基の特異的修飾はそれ自体周知である。例えばスペクトル標識を、チロシル残基に、芳香族ジアゾニウム化

30

40

50

合物またはテトラニトロメタンとの反応によって導入する。N-アセチルイミジズールおよびテトラニトロメタンを使用し、それぞれ0-アセチルチロシル種および3-ニトロ誘導体を形成し得る。カルボキシル側鎖（アスパルチルまたはグルタミル）を、選択的に、カルボジミド（ $R'-N-C-N-R'$ ）例えば1-シクロヘキシル-3-2-モルフォリニル-(4-エチル)カルボジミドまたは1-エチル-3-(4-アゾニア-4,4-ジメチルペンチル)カルボジイミドとの反応によって修飾し得る。

【0039】

さらに、アスパルチルおよびグルタミル残基は、アンモニウムイオンとの反応によってアスパラギニルおよびグルタミニル残基へ変換し得る。グルタミルおよびアスパラギニル残基は、頻繁に、対応するグルタミルおよびアスパルチル残基に脱アミデート化し得る。代替的に、これらの残基を、温和な酸性条件化で脱アミデート化し得る。これらの残基のいずれかの形態は、本発明の範囲内である。

10

【0040】

標的またはスキヤフォールドタンパク質は、3次元構造すなわち、タンパク質のそれぞれの原子についての3次元座標が既知または生成し得る任意のタンパク質であり得る。一般に、これは、X線結晶学的技法、NMR技法、デノボモデリング、相同性モデリング等を使用して決定できる。一般にx線構造を使用するならば、2解像度またはよりよい構造が好ましいが必要ではない。

【0041】

本発明の標的またはスキヤフォールドタンパク質を原核および真核生物、例えばバクテリア（extremeophiles例えばarchebacteria）を含む、真菌、昆虫、魚、植物、および哺乳動物からのものであり得る。好適な哺乳動物は限定しないがげっ歯類（ラット、マウス、ハムスター、ギニアブタ等）、霊長類、家畜（ヒツジ、ヤギ、ブタ、ウシ、ウマ等を含む）およびもとも好ましい実施態様ではヒトからのものである。

20

【0042】

こうして、ここに「標的タンパク質」または「スキヤホルドタンパク質」は、変異体タンパク質または変異体タンパク質のライブラリー、好ましくは変化したコファクター特異性を有するものが、望まれるタンパク質を意味する。当業者によって認識されるように、任意の数の標的タンパク質は、本発明の使用を見出す。「タンパク質」の定義内に具体的に含まれるのは、機能性ドメイン、例えば酵素的ドメイン、結合ドメイン等を含む、既知タンパク質のフラグメントおよびドメイン、および小ドメイン例えばターン、ループ等である。すなわち、タンパク質の部分は同様に使用し得る。加えて、ここで使用するような「タンパク質」は、タンパク質、オリゴペプチドおよびペプチドを含む。加えて、タンパク質変異体すなわち非天然存在タンパク質アナログ構造を使用し得る。

30

【0043】

好適なタンパク質は、限定しないが、工業、製薬、および農業的タンパク質を含む。酵素の好適なクラスは、限定しないが、レダクターゼ、ヒドロラーゼ、例えばプロテアーゼ、カルボヒドラーゼ、リパーゼイソメラーゼ、例えばラセマーゼ、エピメラーゼ、タウトマラーゼ、またはムターゼ、トランスフェラーゼ、キナーゼ、オキシドレダクターゼ、デヒドロゲナーゼ、およびフォスファターゼを含む。好適な酵素は、Seiss-Prot酵素データベースに列挙される。好適なタンパク質バックボーンは、限定しないが、Research Collaboratory for Structural Bioinformatics(RCSB, formerly the Brookhaven National Lab)によってコンパイルおよび供されるタンパク質データベースに見出されるものすべてである。

40

【0044】

具体的には、好ましい標的タンパク質は、レダクターゼ、例えばチオレドキシンレダクターゼ (US Pub. No. 2002/0037303), 2,5-ジケト-D-グルコン酸レダクターゼ(Banta, S, et al., (2002) タンパク質 Eng., 15: 131-140; WO 02/22527; WO 02/29019), グルタチオンレダクターゼ (Mittl, PR, et al. (1993) J. Mol. Biol., 231: 191-5; Mittl & Schulz, (1994) タンパク質 Sci., 3: 799-809; Mittl, PR, et al., (1994) タンパク質 Sci

50

., 3: 1504-14), アルカリヒドロパーオキシドレダクターゼシステム (Wood, ZA, et al., (2001), Biochemistry, 40: 3900-3911), チオレドキシンレダクターゼ-様タンパク質 (Reynolds, CM, et al., (2002) Biochemistry, 41: 1990-2001)を含む。

【0045】

よって本発明は、標的タンパク質のコファクター特異性を変化させるコンピューター的加工方法を指向する。ひとたび標的タンパク質またはスキヤフォールドタンパク質のための座標のセットを入力すると、タンパク質設計サイクルを実行し、望まれるレセプターについての変化した親和性を有する可変タンパク質配列のセットを生成する。ここに「タンパク質設計サイクル」によって限定しないがProtein Design Automation(登録商標) (PDA (登録商標)), 配列予測アルゴリズム (SPA), 種々の力場計算等 (米国特許番号 6,188,965 および6,296,312, 米国出願番号 09/419,351, 09/782,004, 09/927,79, 09/877,695; Rah a, K., et al. (2000) Protein Sci., 9:1106-1119, U.S.S.N. 09/877,695, 2001年6月8日出願, 名称「ApparatusおよびMethod for Designing Proteins and Protein Libraries; 米国出願番号 09/927,790, 60/352,103, および 60/351,937であってすべて引用によりここに全体を含めるものを参照)を含む配列を生産するために使用し得るある数のタンパク質設計アルゴリズムのいずれかを意味する。

10

【0046】

好ましい実施態様では、本発明の方法は、標的タンパク質で出発し、およびコンピューター的加工を使用し、候補または変異体タンパク質または第1配列のセットを生成することに関係する。代替的に、構造に基づく方法、例えばPDA (登録商標) であって以下に詳細に記載するものを使用する。高正確性配列の相対的エネルギーを評価するための他モデルは、引用によりここに含めるWarshel, Computer Modeling of Chemical Reactions in EnzymesおよびSolutions, Wiley & Sons, New York, (1991)を含む。

20

【0047】

同様に、分子力学計算を使用し、突然変異体配列スコアを個々に計算し、およびランク整列リストをコンパイルすることによって配列をコンピューター的にスクリーンすることができる。

【0048】

好ましい実施態様では、残基対ペナントを使用し、配列をスコア化することができる (M iyazawa et al., Macromolecules 18(3):534-552 (1985), 引用によりここに含める) during computational screening.

30

【0049】

好ましい実施態様では、配列プロファイルスコア (Bowie et al., Science 253(5016):164-70 (1991), 引用によりここに含める) および/または平均力の潜在性 (Hendlich et al., J. Mol. Biol. 216(1):167-180 (1990), また引用によりここに含める) はまた、配列をスコア化するために計算し得る。これらの方法は、配列と3Dタンパク質構造の間のマッチを評価し、ゆえにタンパク質構造への忠実性についてスクリーンするために作用できる。配列をランクするため種々のスコア化機能を使用することによって、配列スペースの種々の領域を、コンピューター的スクリーンにおいてサンプリングできる。

【0050】

さらに、スコア化機能を使用し、タンパク質の金属またはコファクター結合部位を創出する配列についてスクリーンできる。(Hellenga, Fold Des. 3(1): R1-8 (1998), 引用によりここに含める)。同様に、機能をスコアすることを使用し、タンパク質にジスルフィドを創出する配列についてスクリーンできる。これらの潜在性は新しい構造モチーフを導入得るタンパク質構造を具体的に修飾するのを試みる。

40

【0051】

好ましい実施態様では、配列および/または構造アラインメントプログラムを使用し、発明の変異体タンパク質を生成し得る。当業界で知られるように、ある数の配列に基づくアラインメントプログラム、例えば、Smith-Waterman searches, Needleman-Wunsch, Double Affine Smith-Waterman, frame search, Gribskov/GCG profile search, Gribskov/GCG

50

profile scan, profile frame search, Bucher generalized profiles, Hidden Markov models, Hframe, Double Frame, Blast, Psi-Blast, Clustal, および GeneWise を含むものがある。

【0052】

配列の源は広く変化し、そして1または2以上の既知のデータベースであって限定しないが SCOP (Hubbard, et al., *Nucleic Acids Res* 27(1):254-256. (1999)); PFAM (Bateman, et al., *Nucleic Acids Res* 27(1):260-262. (1999)); VAST (Gibrat, et al., *Curr Opin Struct Biol* 6(3):377-385. (1996)); CATH (Orengo, et al., *Structure* 5(8):1093-1108. (1997)); PhD Predictor (<http://www.embl-heidelberg.de/predictタンパク質/predictタンパク質.html>); Prosite (Hofmann, et al., *Nucleic Acids Res* 27(1):215-219. (1999)); PIR (<http://www.mips.biochem.mpg.de/proj/protseqdb/>); GenBank (<http://www.ncbi.nlm.nih.gov/>); PDB (www.rcsb.org) および BIND (Bader, et al., *Nucleic Acids Res* 29(1):242-245. (2001)) を含むものからの配列をとることを含む。

【0053】

加えて、これらのデータベースからの配列を、連続分析または遺伝子予測の対象とすることができる; Wheeler, et al., *Nucleic Acids Res* 28(1):10-14. (2000) および Burge および Karlin, *J Mol Biol* 268(1):78-94. (1997) 参照。

【0054】

当業界で知られるように、使用できる配列アラインメント方法論のある数がある。例えば、配列相同性に基づくアラインメント方法を使用し、標的構造に関連するタンパク質の配列アラインメントを創出できる (Altschul et al., *J. Mol. Biol.* 215(3):403-410 (1990), Altschul et al., *Nucleic Acids Res.* 25:3389-3402 (1997), both 引用によりここに含める)。これらの配列アラインメントを次いで試験し、観察配列変異を決定する。これらの配列変異を集計し、変異体タンパク質のあるセットを定義する。

【0055】

配列に基づくアラインメントを、種々の方法で 사용할 ことができる。例えば関連タンパク質のある数を、当業界で既知のようにアラインメントし、「可変」および「保存」残基を定義する。これらの結果を使用し、以下に概説の確率表を生成することができる。同様に、これらの配列変異を、作表し、そして二次ライブラリーを以下に定義のようにそれらから定義する。代替的に、もたらされた配列変異を使用し、コンピューターのスクリーニング中のそれぞれの位置で考えられるアミノ酸を定義することができる。他の変異は、配列アラインメントで存在するアミノ酸についてのスコアをバイアスさせ、それによってそれらがコンピューターのスクリーニング中に見出される尤度を増加させるが、なお他のアミノ酸の考察を可能とする。このバイアスは変異体タンパク質のフォーカス化ライブラリーという結果となるが、アラインメントで見出されない考察アミノ酸から排除しない。加えて、他の型のある数は、導入し得る。例えば、多様性をフォーカスし得る; すなわち、「保存」残基を選択し、変化させ、タンパク質上の多様性を強化し、こうして配列スペースのより大きい部分をサンプリングする。代替的に、ファミリーメンバー間の高可変性の位置 (すなわち低保存) は、アミノ酸のすべてまたはサブセットを使用してランダム化できる。同様に、ポジショナルアウトライアーまたは側鎖アウトライアーを排除し得る。

【0056】

同様に、構造に基づくタンパク質の構造アラインメントをなし、配列アラインメントを生成し得る。既知の広範囲のそのような構造アラインメントプログラムがある。例えば VAST from the NCBI (<http://www.ncbi.nlm.nih.gov:80/Structure/VAST/vast.shtml>); SSAP (Orengo and Taylor, *Methods Enzymol* 266(617-635 (1996)) SARF2 (Alexandrov, *Protein Eng* 9(9):727-732. (1996)) CE (Shindyalov and Bourne, *Protein Eng* 11(9):739-747. (1998)); (Orengo et al., *Structure* 5(8):1093-108 (1997); Dali (Holm et al., *Nucleic Acid Res.* 26(1):316-9 (1998) であって、すべて引用によりここに含めるもの参照)。これらの配列アラインメントを次いで試験し、観察配列変異を決定できる。ライブラリーを、配列柄の二次構造を予測し、次いで予測される二次構造と矛盾しない配列を選

択することによって生成できる。ある数の二次構造予測方法、例えばヘリックス-コイル遷移理論(MunozおよびSerrano, Biopolymers 41:495, 1997), 神経ネットワーク, 局所構造アラインメントおよびその他 (例えば、Selbig et al., Bioinformatics 15:1039-46, 1999参照)がある。

【0057】

同様に、前概説の通り、限定しないが配列プロファイリング[Bowie and Eisenberg, Science 253(5016):164-70, (1991)], ロタマーライブラリー (rotamer library) selection [Dahiyat and Mayo, Protein Sci. 5(5):895-903 (1996); Dahiyat and Mayo, Science 278(5335):82-7 (1997); Desjarlais and Handel, Protein Science 4:2006-2018 (1995); Harbury et al, Proc. Natl. Acad. Sci. U.S.A. 92(18):8408-8412 (1995); Kono et al., Proteins: Structure, Function and Genetics 19:244-255 (1994); Hellinga and Richards, Proc. Natl. Acad. Sci. U.S.A. 91:5803-5807 (1994)]; および残基対ポテンシャル[Jones, Protein Science 3: 567-574, (1994)]; PROSA [Heindlich et al., J. Mol. Biol. 216:167-180 (1990)]; THREADER [Jones et al., Nature 358:86-89 (1992)], および他の逆フォールディング方法例えばSimons et al. [Proteins, 34:535-543, (1999)], LevittおよびGerstein [Proc. Natl. Acad. Sci. U.S.A., 95:5913-5920, (1998)], Godzik and Skolnick [Proc. Natl. Acad. Sci. U.S.A., 89:12098-102, (1992)], Godzik et al. [J. Mol. Biol. 227:227-38, (1992)] によって記載されるものおよび2プロファイリング法 [Gribskov et al. Proc. Natl. Acad. Sci. U.S.A. 84:4355-4358 (1987) およびFischer and Eisenberg, Protein Sci. 5:947-955 (1996), Rice and Eisenberg J. Mol. Biol. 267:1026-1038 (1997)]であってすべて 引用によりここに含めるものを含む他のコンピューター的方法が、既知である。

【0058】

加えて、他のコンピューター的方法、例えば Koehl and Levitt (J. Mol. Biol. 293:1161-1181 (1999); J. Mol. Biol. 293:1183-1193 (1999); 引用によりここに含める) によって記載されるものを使用し、改善された特性および機能のための実験スクリーニングで使用するのより小さい二次ライブラリーを生成するため所望により次いで使用することのできる変異体ライブラリーを創出し得る。加えて、力場計算 (force-field calculations) 例えばSCMFであってSCMFについてまた使用できるものに基づくコンピューター的方法がある。Delarue et al. Pac. Symp. Biocomput. 109-21 (1997); Koehl et al., J. Mol. Biol. 239:249-75 (1994); Koehl et al., Nat. Struct. Biol. 2:163-70 (1995); Koehl et al., Curr. Opin. Struct. Biol. 6:222-6 (1996); Koehl et al., J. Mol. Biol. 293:1183-93 (1999); Koehl et al., J. Mol. Biol. 293:1161-81 (1999); Lee J., J. Mol. Biol. 236:918-39 (1994); およびVasquez Biopolymers 36:53-70 (1995)であってすべて引用によりここに含めるもの参照。他の力場計算 (forcefield calculations) であってコンピューター的方法内で配列のコンホメーションを最適化するため、または前概説のようにデノボ最適化配列を生成するため使用できるものは、限定しないが、OPLS-AA [Jorgensen et al., J. Am. Chem. Soc. 118:11225-11236 (1996); Jorgensen, W.L.; BOSS, Version 4.1; Yale University: New Haven, CT (1999)]; OPLS [Jorgensen et al., J. Am. Chem. Soc. 110:1657ff (1988); Jorgensen et al., J. Am. Chem. Soc. 112:4768ff (1990)]; UNRES (United Residue Forcefield; Liwo et al., Protein Science 2:1697-1714 (1993); Liwo et al., Protein Science 2:1715-1731 (1993); Liwo et al., J. Comp. Chem. 18:849-873 (1997); Liwo et al., J. Comp. Chem. 18:874-884 (1997); Liwo et al., J. Comp. Chem. 19:259-276 (1998); Forcefield for Protein Structure Prediction (Liwo et al., Proc. Natl. Acad. Sci. U.S.A. 96:5482-5485 (1999)]; ECEPP/3 [Liwo et al., J. Protein Chem. 13(4):375-80 (1994)]; AMBER 1.1 force field (Weiner et al., J. Am. Chem. Soc. 106:765-784); AMBER 3.0 force field [U.C. Singh et al., Proc. Natl. Acad. Sci. U.S.A. 82:755-759 (1985)]; CHARMM and CHARMM22 (Brooks et al., J. Comp. Chem. 4:187-217); cvff3.0 [Dauber-Osguthorpe et al., Proteins: Structure, Function and Genetics, 4:31-47 (1988)]; cff91 (Maple et al., J. C

omp. Chem. 15:162-182)を含み; またDISCOVER (cvff and cff91)および AMBER force-fieldsをINSIGHT molecular modeling package (Biosym/MSI, San Diego California)で使用し得、そして HARMMをQUANTA molecular modeling package (Biosym/MSI, San Diego California)で使用し得、すべて引用によりここに含める。実際、以下に概説のようにこれらのforce-field方法を使用し、変異体 TR ライブラリーを直接的に生成し得;これらの方法を使用して、追加的ライブラリーを直接的に生成する確率表を生成できる。

【0059】

好ましい実施態様では、Protein Design Automation (登録商標) (PDA (登録商標)) を使用し、すべて引用によりここに含める米国特許番号 6,188,965および6,296,312に記載されるようなそれぞれのアミノ酸位置についての定義されたエネルギー状態を含む可変タンパク質配列を生成する。簡単には、PDA (登録商標) を以下のように記載できる。既知のタンパク質構造を出発点として使用する。最適化すべき残基を次いで同定し、これは完全配列またはそのサブセットであり得る。変化すべき任意の位置の側鎖を次いで除去する。得られた構造は、タンパク質バックボーンからなり、そして残りの側鎖はテンプレートと呼ばれる。それぞれの可変残基位置を次いで好ましくは、コア残基、表面残基、または境界残基として分類し;それぞれの分類は、位置についての可能なアミノ酸残基のサブセットを定義する(例えば、コア残基は一般に疎水性残基のセットから選択され、表面残基は一般に、親水性残基から選択され、および境界残基はいずれかから選択される)。それぞれのアミノ酸は、すべての可能な、ロタマーと呼ばれるそれぞれの側鎖コンホマーの明確なセットによって表すことができる。こうして、バックボーンについて最適配列で生存するために、ロタマーのすべての可能な配列を、スクリーニングしなければならない、ここでそれぞれのバックボーン位置は、すべてのその可能なロタマー性状態のそれぞれのアミノ酸、またはアミノ酸のサブセット、こうしてロタマーのサブセットによって占有されることができる。

【0060】

相互作用の2セットを次いで、すべての位置のそれぞれのロタマーについて計算する: バックボーンの全部または一部を有する、ロタマー側鎖の相互作用(「シングル」エネルギー、またロタマー/ロタマーエネルギーと呼ばれる)、およびすべての他の位置または他の位置のサブセットの、すべての他の可能なロタマーを有するロタマー側鎖の相互作用(「ダブル」エネルギー、またロタマー/ロタマーエネルギーと呼ばれる)。これらの相互作用のそれぞれのエネルギーを、種々のスコアリング関数の使用によって計算し、これはファンデルワールス力、水素結合のエネルギー、二次構造的向のエネルギー、表面面積溶媒和化および静電気を含む。こうして、それぞれのロタマー相互作用の総エネルギー、バックボーンおよび他のロタマーの両方を計算し、そしてマトリクス形態に貯蔵する。

【0061】

ロタマーセットの明確な性質は、試験すべきロタマー配列の数の単純な計算を可能とする。位置あたり m 個の可能なロタマーを有する長さ n のバックボーンは、 mn 可能ロタマー配列を有し、配列長で指数関数的に成長しかつリアルタイムで扱いにくいまたは不可能な計算にさせる。よって、この組み合わせサーチ問題を解決するために、「Dead End Elimination」(DEE)計算を実施する。DEE計算は、もし第1のロタマーの最悪の総相互作用がなお第2ロタマーの最良の総相互作用よりもよりよいならば、そのとき第2ロタマーは、グローバル最適ソリューションの一部であることはできないという事実に基づく。すべてのロタマーのエネルギーがすでに計算されたので、DEEアプローチは、ロタマーを試験および排除するための配列長に渡る和を要するのみであり、これは計算をかなりスピードアップする。DEEはロタマーの対の比較、またはロタマーの組み合わせに戻る事ができ、これは最終的に、グローバル最適エネルギーを表す単一配列の決定という結果となる。

【0062】

グローバルソリューションが見出されると、Monte Carloサーチをなし得、DEEソリューションの近所の配列のランク整列リストを生成し得る。DEEソリューションで出発す

10

20

30

40

50

ると、ランダム位置は、他のロタマーに変化し、そして新しい配列エネルギーを計算する。もし新しい配列が、許容の基準に合致するならば、それを他のジャンプのための出発点として使用する。前決定数のジャンプ後、配列のランク整列リストを生成する。Monte Carloサーチングは、サンプリング技法であり、グローバル最小周辺の配列スペースを探索し、または配列スペース中の新しいローカル最小距離を見出す。より追加的に以下に概説するように、使用できる他のサンプリング技法があり、Boltzmanサンプリング、遺伝的アルゴリズム技法、およびシミュレート性アニーリングを含む。加えて、すべてのサンプリング技法に浮いて、可能であるジャンプの種類は、変化できる（例えば、ランダム残基へのランダムジャンプ、バイアスジャンプ（野生型にまたはそれから、例えば）バイアス残基へのジャンプ（類似残基にまたはそれから、例えば）等）。同様に、すべてのサンプリング技法について、サンプリングジャンプが許容されるかの許容される基準は、変化できる。

10

【0063】

米国出願番号09/127926に概説されるようにタンパク質バックボーン（（天然存在タンパク質のために）窒素、カルボニル炭素、 α -炭素、および α 炭素から β 炭素へのベクターの方向と一緒にカルボニル酸素を含む）を、コンピューター的分析に先立ち、超二次構造パラメーターと呼ばれるパラメーターのセットを変化させることによって変化させ得る。

【0064】

タンパク質バックボーンを生成すると（前概説のような変化）、およびコンピューター入力すると、明示的水素を、もし構造内に含まれないなら追加する（例えばもし構造がx線結晶学によって生成されたならば、水素を追加すべきである）。水素追加後に、構造のエネルギー最小化を稼動し、水素並びに他の原子、結合エネルギーおよび結合長を弛緩する。好ましい実施態様では、これは、原子座標位置のコンジュゲート勾配最小化（Mayo et al., J. Phys. Chem. 94:88997(1990)）のある数の工程によってなし、静電気のないDreiding力場を最小化する。約10ないし約250工程が好ましく、約50が最も好ましい。

20

【0065】

少なくとも1可変残基位置でタンパク質バックボーン構造を含む。当業界で知られるように、タンパク質の残基またはアミノ酸は、一般に、タンパク質のN末端で開始し逐次的に数えられる。こうしてそのN末端のメチオニンを有するタンパク質は、アミノ酸または残基位置1のメチオニンを有するといわれ次の残基は2、3、4等という。それぞれの位置で、野生型（すなわち天然存在）タンパク質は、任意の数のロタマーの少なくとも20アミノ酸の1つを有し得る。「可変残基位置」によって、ここで、設計されるべきタンパク質のアミノ酸位置を意味し、これは特異的残基またはロタマー、一般に野生型残基またはロタマーとして設計方法において固定化されない。

30

【0066】

好ましい実施態様では、タンパク質の残基位置のすべては可変である。すなわち、すべてのアミノ酸側鎖を、本発明の方法において変化し得る。これは特に、より小さいタンパク質について望ましく、しかし本方法は、より大きいタンパク質の設計をまた可能とする。このやり方で設計し得るタンパク質の長さに理論的な制限はないが、実際的なコンピューターの制限がある。

40

【0067】

代替的好ましい実施態様では、タンパク質の残基位置のいくつかのみが可変であり、そして残りは「固定化」であり、すなわち、これらはあるセットのコンホメーションであるとして3次元構造で同定される。いくつかの実施態様では、固定化位置は、そのもとのコンホメーションで残る（これは使用されるロタマーライブラリーの特異的ロタマーと関連し得ない）。代替的に、残基を非野生型残基として固定化し得；例えば、既知部位特異的突然変異誘発技法が、特定残基が望ましいことを示したときに（例えば、タンパク質溶解位置または酵素の基質特異性を変化させる）、残基を特定アミノ酸として固定化し得る。代替的に、本発明の方法を使用し、デノボ突然変異を評価し得る。以下に議論の通りであ

50

る。代替的好ましい実施態様では、固定化位置は「フロート性」であり得、その位置のアミノ酸を固定化するが、そのアミノ酸の異なるロタマーを試験する。この実施態様では、可変残基は、少なくとも1つであり得、任意のばあい、残基の総数の0.1%ないし99.9%である。こうして、例えば、わずか2、3（または1）残基、またはほとんどの残基、その間のすべての可能性がありつつ変化させることが可能であり得る。

【0068】

好ましい実施態様では、固定化できる残基は、限定しないが、構造的または生物学的機能性残基を含み；代替的に、生物学的機能性残基は、特異的に固定化し得ない。例えば、生物学的活性について重要である、例えば、酵素の活性部位、酵素の基質結合部位、結合パートナーについての結合部位（リガンド／レセプター、抗原／抗体等）、ホスホリル化またはグリコシル化部位であって生物学的機能に重要であるもの、または構造的な重要残基、例えばジスルフィド架橋、金属結合部位、重要水素結合残基、バックボーンコンホメーションについて重要な残基、例えばプロリンまたはグリシン、パッキング相互作用に重要な残基、等を形成することが既知である残基は、すべてコンホメーション中に、または単一ロタマーまたは「フロート化」であり得る。

10

【0069】

同様に、可変残基として選択され得る残基は、望ましくない生物学的属性、例えばタンパク質溶解性分解、ダイマー化または凝集部位、グリコシル化部位であって免疫応答につながり得るもの、望まれない結合活性、望まれないアロステリー、結合の保存を有するが望まれない酵素活性への感受性、等を付与するものであり得る。

20

【0070】

好ましい実施態様では、それぞれの可変位置を、コア、表面または境界残基位置として分類するが、しかし以下に説明のようにある場合には、可変位置をグリシンにセットし得、バックボーン系統を最小化する。加えて、ここに概説のように、残基は分類を要せず、これらを変数として選択肢、そしてアミノ酸の任意のセットを使用し得る。コア、表面および境界位置の任意の組み合わせを利用できる：コア、表面および境界残基；コアおよび表面残基；コアおよび境界残基、および表面および境界残基、並びにコア残基単体、表面残基単体、または境界残基単体。

【0071】

コア、表面または境界としての残基位置の分類は、いくつかのやり方でなし得、当業者によって認識される通りである。好ましい実施態様では、分類を、側鎖を含むもとのタンパク質バックボーン構造の視覚的スキャンによってなし、そしてタンパク質のモデリングの技術での当業者の主観的評価に基づく分類を割り当てる。代替的に、好ましい実施態様は、テンプレートC α 原子のみを使用してコンピューター計算された溶媒接近可能表面に対して、C α -C β ベクターの配向の評価を利用し、60/061,097、60/043,464、60/054,678、09/127,926およびPCT US98/07254に記載される通りである。代替的に、表面面積計算をなすことができる。

30

【0072】

ひとたびそれぞれの可変位置を所望により、コア、表面または境界として分類すると、アミノ酸側鎖のセットおよびこうしてロタマーのセットを、それぞれの位置に割り当てる。すなわち、可能なアミノ酸側鎖のセットであって、任意の特定の位置でプログラムが考えることができるものを、選択する。逐次的に、ひとたび可能なアミノ酸そくさが選択されると、特定の位置で評価させるであろうロタマーのセットを決定できる。こうして、コア残基は一般に、アラニン、バリン、イソロイシン、ロイシン、フェニルアラニン、チロシン、トリプトファン、およびメチオニンからなる親水性残基の群から選択される（いくつかの実施態様では、以下に記載のようなファンデルワールススコアリング関数のoscaling factorが低いとき、メチオニンをセットから除去する）、そしてそれぞれのコア位置についてのロタマーセットは潜在的に、これらの8アミノ酸側鎖についてのロタマーを含む（もしバックボーン独立ライブラリーを使用するならば、すべてのロタマーそしてもしロタマー依存性バックボーンを使用するならばサブセット）。同様に、表面位置は一般に、

40

50

アラニン、セリン、アスパラギン酸、アルパラギン、グルタミン、グルタミン酸、アルギニン、リジンおよびヒスチジンからなる親水性残基の群より選択する。それぞれの表面位置についてのロタマーえっとは、こうして、これらの10残基についてのロタマーを含む。最終的に、境界位置は一般に、アラニン、セリン、トレオニン、アスパラギン酸、アスパラギン、グルタミン、グルタミン酸、アルギニン、リジン、ヒスチジン、バリン、イソロイシン、ロイシン、フェニルアラニン、チロシン、トリプトファンおよびメチオニンから選択される。それぞれの境界位置についてのロタマーセットは、こうして潜在的に、これらの17残基についてのすべてのロタマーを含む（システイン、グリシン、およびプロリンは使用しないと仮定するが、これらは可能である）。追加的に、いくつかの好ましい実施態様では、18の天然存在アミノ酸のセット（すべてシステインとプロリンを除き、これらは特に破壊的であると知られる）を使用する。 10

【0073】

こうして、当業者に認識されるように、残基位置を分類するコンピューター的利益が存在し、というのは計算の数を減少させるからである。またコア、境界および表面残基が、前記のものから変化されるような状況があってもよいことが注意されるべきであり；例えば幾つかの状況下で、1または2以上のアミノ酸が、もたらされるアミノ酸のセットから付加または差し引きされる。例えば、ダイマー化、またはマルチマー化する、またはリガンド結合部位を有する、いくつかのタンパク質は、親水性表面残基等を含み得る。加えて、ヘリックス「キャッピング」をもたらし残基、またはヘリックス双極子との優先的相互作用は、もたらされる残基のセットから差し引きされ得る。アミノ酸群のこの修飾は、残基による残基基本に基づいてなす。 20

【0074】

好ましい実施態様では、プロリン、システインおよびグリシンは、可能なアミノ酸側鎖のリストに含まれず、そしてこうして、これらの側鎖についてのロタマーは、使用しない。しかし、好ましい実施態様では、可変残基位置が0より大きいφ角を有するとき（すなわち、2平面角が、（1）先行アミノ酸のカルボニル炭素、（2）現在の残基の窒素原子（3）現在の残基のα炭素および（4）現在の残基のカルボニル炭素によって定義される）、位置は、バックボーン緊張を最小化するグリシンにセットされる。

【0075】

潜在的ロタマーの群がそれぞれの可変位置に割り当てられると、米国出願番号09/127926 およびPCTUS98/07254に概説のように、加工は進行する。加工工程は、最適タンパク質配列を生成するために、ロタマーの、互いとおよびタンパク質バックボーンとの相互作用を分析することを必要とする。簡単には加工は理想的には、ロタマーの、バックボーンそれ自体または他のロタマーとの相互作用のエネルギーを計算するためにある数のスコア化関数の使用を含む。好ましいPDAスコア関数は限定しないが、ファンデルワールスポテンシャルスコア化関数、水素結合ポテンシャルスコア化関数、原子溶媒和化スコア化関数、二次構造特性スコア化関数および静電的スコア化関数を含む。さらに以下に記載するように、少なくとも1のスコア化関数を使用し、それぞれの位置をスコア化するが、スコア化関数は、位置分類または他の考察、例えばα双極子との優先される相互作用に基づいて異なり得る。以下に概説するように、計算で使用される総エネルギーは、特定の位置で使用されるそれぞれのスコア化関数のエネルギーの合計であり、等式1で一般的に示すとおりである： 30 40

等式 1

$$E_{total} = nE_{vdw} + nE_{as} + nE_{h-bonding} + nE_{ss} + nE_{elec}$$

【0076】

等式1において、総エネルギーは、ファンデルワールスポテンシャルのエネルギー(E_{vdw})、原子溶媒和化のエネルギー(E_{as})、水素結合のエネルギー($E_{h-bonding}$)、二次構造のエネルギー(E_{ss})および静電相互作用のエネルギー(E_{elec})の合計である。用語nは0または1であり、該用語が特定の残基位置について考察されるべきであるかに依存するものである。

【0077】

米国出願番号 60/061,097, 60/043,464, 60/054,678, 09/127,926およびPCT US98/07254に概説されるように、単体または組み合わせの、これらのスコア化関数の任意の組み合わせを使用し得る。ひとたび使用すべきスコア化をそれぞれの可変位置について同定すると、コンピューター的分析における好ましい第1工程は、タンパク質の残りの全部または一部とのかそれぞれの可能なロタマーの相互作用の決定を含む。すなわち、バックボーンまたは他のロタマーとのそれぞれの可変残基位置のそれぞれの可能なロタマーの、1または2以上のスコア化関数によって測定されるような相互作用のエネルギーを、計算する。好ましい実施態様では、タンパク質の全体の残りとの、それぞれのロタマーの相互作用、すなわち全体のテンプレートおよびすべての他のロタマーの両方がなされる。しかし、前概説のとおり、タンパク質の一部、例えばより大きいタンパク質の1ドメインをモデル化することのみ可能であり、こうして幾つかの場合には、タンパク質のすべては考察する必要はない。ここで使用する用語「一部」は、タンパク質についてそのタンパク質のフラグメントをいう。このフラグメントは、10アミノ酸残基のサイズから、完全アミノ酸配列マイナス1アミノ酸までの範囲であり得る。よって、ここで使用する用語「一部」は、核酸について、その核酸の1フラグメントを意味する。このフラグメントは、10ヌクレオチドのサイズから、完全核酸配列マイナス1ヌクレオチドまでの範囲であり得る。

10

【0078】

好ましい実施態様では、コンピューター的加工の第一工程は、すべての位置のそれぞれのロタマーについての相互作用の2セットを計算することをによってなす：その位置が変異またはフロートであっても、ロタマー側鎖の、テンプレートまたはバックボーンとの相互作用（「シングル」エネルギー）およびすべての他の位置のすべての他の可能なロタマーとのロタマー側鎖の相互作用（「ダブル」エネルギー）。この場合のバックボーンはタンパク質構造バックボーンの原子並びに任意の固定化された残基の原子の両方を含み、ここで該固定化残基は、アミノ酸の特定のコンホメーションとして定義される。

20

【0079】

こうして、「シングル」（ロタマー／テンプレート）エネルギーは、いくつかまたはすべてのスコア化関数を使用して、バックボーンとの、すべての可変位置の残基位置のすべての可能なロタマーの相互作用について計算する。こうして、水素結合スコア化関数のために、ロタマーのすべての水素結合原子およびバックボーンのすべての水素結合原子を評価し、そして E_{HB} をすべての可変位置のそれぞれの可能なロタマーについて計算する。同様に、ファンデルワールススコア化関数について、ロタマーのすべての原子を、テンプレートのすべての原子と比較し（一般に、それ自身の残基のバックボーン原子を除外する）、そして E_{vdW} を、すべての可変残基位置のそれぞれの可能なロタマーについて計算する。加えて、もし原子が3結合またはより少なく接続されるならば、ファンデルワールスエネルギーは計算されない。原子溶媒和化スコア化関数のために、ロタマーの表面を、テンプレートの表面に対して測定し、そしてすべての可変残基位置のそれぞれの可能なロタマーについての E_{as} を計算する。二次構造的向スコア化関数をまた、シングルエネルギーとして考察し、こうして、総シングルエネルギーは、 E_{ss} 期間を含み得る。当業者によって認識されるように、これらのエネルギー期間の多くは、ロタマーとテンプレート位置の間の物理的距離に依存してゼロに近くなり；すなわち、2部分がさらに離れるにつれ、エネルギーはより低い。

30

40

【0080】

「ダブル」エネルギー（ロタマー／ロタマー）の計算のために、それぞれの可能なロタマーの相互作用エネルギーを、すべての他の可変残基位置のすべての可能なロタマーと比較する。こうして、「ダブル」エネルギーを、幾つかまたはすべてのスコア化関数を使用して、すべての他の可変残基位置すべての可能なロタマーとの、すべての可変残基位置のすべての可能なロタマーの相互作用のために計算する。こうして、水素結合スコア化関数のために、第1ロタマーのすべての水素結合原子およびすべての可能な第2ロタマーのすべての水素結合原子はを評価し、そして E_{HB} を、任意の2可変位置についてのそれぞれの可

50

能なロタマー対について計算する。同様に、ファンデルワールスコア化関数のために、第1ロタマーのすべての原子を、すべての可能な第2ロタマーのすべての原子と比較し、そして E_{vdw} を、すべての2可変残基位置のそれぞれの可能なロタマー対について計算する。原子溶媒和化スコア化関数のために、第1ロタマーの表面は、すべての可能な第2ロタマーの表面に対して測定し、そしてすべての2可変残基位置のそれぞれの可能なロタマー対についてのEasを計算する。二次構造スコア化関数は、「ダブル」エネルギーとして稼動する必要はなく、「シングル」エネルギーの構成要素として考察されたとおりである。当業界で認識されるように、これらのダブルエネルギー期間の多くは、第1ロタマーと第2ロタマーの間の物理的距離に依存して、近いないしゼロであり；すなわち2部分がよりさらに離れるにつれ、エネルギーはより低い。

10

【0081】

好ましい実施態様では、配列予測アルゴリズム (SPA) を使用し、両方引用によりここに含める、Raha, K., et al. (2000) タンパク質 Sci., 9:1106-1119, 2001年6月8日出願の米国出願09/877,695名称「Apparatus and Method for Designing Proteins and Protein Libraries」に記載されたようにそれぞれのアミノ酸位置について定義されたエネルギー状態を含む可変タンパク質配列を生成する。

【0082】

好ましい実施態様では、力場計算例えばSCMFを使用し、それぞれのアミノ酸位置について定義されるエネルギー状態を含む可変タンパク質配列を生成できる。SCMFのために、すべて引用によりここに含める、Delarue et al., Pac. Symp. Biocomput. 109-21 (1997); Koehl et al., J. Mol. Biol. 239:249 (1994); Koehl et al., Nat. Struct. Biol. 2:163 (1995); Koehl et al., Curr. Opin. Struct. Biol. 6:222 (1996); Koehl et al., J. Mol. Biol. 293:1183 (1999); Koehl et al., J. Mol. Biol. 293:1161 (1999); Lee J. Mol. Biol. 236:918 (1994);およびVasquez Biopolymers 36:53-70 (1995)参照。コンピューター的方法内で配列のコンホメーションを最適化するために、またはここに概説のようなデノボ最適化配列を生成するため使用できる他の力場計算は、限定しないがすべて引用によりここに含める、OPLS-AA (Jorgensen, et al., J. Am. Chem. Soc. (1996), v 118, pp 11225-11236; Jorgensen, W.L.; BOSS, Version 4.1; Yale University: New Haven, CT (1999)); OPLS (Jorgensen, et al., J. Am. Chem. Soc. (1988), v 110, pp 1657ff; Jorgensen, et al., J Am. Chem. Soc. (1990), v 112, pp 4768ff); UNRES (United 残基 Forcefield; Liwo, et al., Protein Science (1993), v 2, pp1697-1714; Liwo, et al., Protein Science (1993), v 2, pp1715-1731; Liwo, et al., J. Comp. Chem. (1997), v 18, pp849-873; Liwo, et al., J. Comp. Chem. (1997), v 18, pp874-884; Liwo, et al., J. Comp. Chem. (1998), v 19, pp259-276; Forcefield for Protein Structure Prediction (Liwo, et al., Proc. Natl. Acad. Sci. USA (1999), v 96, pp5482-5485); ECEPP/3 (Liwo et al., J Protein Chem 1994 May 13(4):375-80); AMBER 1.1 force field (Weiner, et al., J. Am. Chem. Soc. v106, pp765-784); AMBER 3.0 force field (U.C. Singh et al., Proc. Natl. Acad. Sci. USA. 82:755-759); CHARMM and CHARMM22 (Brooks, et al., J. Comp. Chem. v4, pp 187-217); cvff3.0 (Dauber-Osguthorpe, et al., (1988) Proteins: Structure, Function and Genetics, v4, pp31-47); cff91 (Maple, et al., J. Comp. Chem. v15, 162-182)を含み;またDISCOVER (cvff and cff91)およびAMBER forcefieldsをINSIGHT molecular modeling package (Biosym/MSI, San Diego California)で使用しそしてHARMMをQUANTA molecular modeling package (Biosym/MSI, San Diego California)で使用する。実際、以下に概説のように、力場方法を使用し、二次的ライブラリーを直接的に生成し；すなわち一次的ライブラリーを生成せず；むしろ、これらの方法を使用し、二次的ライブラリーを直接的に生成する確率表を、例えばSCMF計算中にこれあの力場を使用することによって生成できる。

20

30

40

【0083】

ひとたびシングルおよびダブルエネルギーを計算し、蓄積すると、コンピューターの加工の次の工程が起こり得る。米国出願09/127926およびPCTUS98/07254に概説されるように、

50

好ましい実施態様は、Dead End Elimination (DEE)工程、および好ましくはMonte Carlo工程を利用する。

【0084】

広く調べられる、PDA（登録商標）は、出力（例えば一次的ライブラリー）を変異させ変化させ得る3つの構成要素を有する：過程で使用されるスコアー化関数；フィルター化技法、およびサンプリング技法。

【0085】

好ましい実施態様では、スコアー化関数を、変化させ得る。好ましい実施態様では、前概説スコアー化関数を種々の方法でバイアスまたは秤量し得る。例えば参照配列または配列のファミリーに向かってまたはそれからのバイアスをなし得る；例えば、野生型またはホモログ残基に向かうバイアスを使用し得る。同様に、その完全タンパク質またはフラグメントをバイアスし得る；例えば化性部位を野生型残基に向かって偏らせ、または特定の望まれる物理的特性に向かうドメイン残基をなすことができる。さらに、増加エネルギーに向かうまたはそれに対するバイアスを生成できる。追加的スコアー化関数は、限定しないが静電的ポテンシャル勾配または疎水性勾配を適用すること、基質を付加すること、または計算へのパートナーを結合すること、または望まれる変化または疎水性へむかってバイアスさせることを含む。

10

【0086】

加えて、代替的实施態様では、使用し得る種々の追加的スコアー化関数がある。追加的スコアー化関数は、限定しないが、ねじれポテンシャルまたは残基対ポテンシャル、または残基エントロピーポテンシャルを含む。そのような追加的スコアー化関数を単独でまたはそれを当所にスコアー化した後にライブラリーを加工するための関数として使用し得る。例えば、MHC（Major Histocompatibility Complex）へのペプチドの結合についてのデータから由来する種々の関数を使用し、MHCに潜在的に結合できる配列、すなわち潜在的に免疫発生性配列を含むタンパク質をするために再スコアー化できる。

20

【0087】

好ましい実施態様では、種々のフィルター化技法をなし得、限定しないがDEEおよびその関連カウンターパートを含む。追加的フィルター化技法は、限定しないが、最適配列を見出すためのbranch-and-bound技法（Gordon and Mayo, Structure Fold. Des. 7:1089-98, 1999）および配列のexhaustive enumerationを含む。しかし、いくつかの技法を任意のフィルター化技法なくしてなし得；例えばサンプリング技法を使用し、フィルター化なくして良好な配列を見出すことができる。

30

【0088】

当業者に認識されるように、ひとたび配列または配列のセットを生成すると、Monte Carlo方法に加えて、またはMonte Carloサーチの代わりに種々の配列スペースサンプリング方法をなし得る。すなわち、ひとたび配列または配列のセットを生成すると、好ましい方法は、試験のために追加的、関連配列の生成を可能とするサンプリング技法を利用する。

【0089】

これらのサンプリング方法は、アミノ酸置換、挿入または欠失、または1または2以上の配列の組換えの使用を含むことができる。ここに概説のように、好ましい実施態様は、Monte Carloサーチを利用し、これは一連のバイアス、システマティックまたはランダムジャンプである。しかし、使用できる他のサンプリング技法があり、Boltzmanサンプリング、遺伝的アルゴリズム技法およびシミュレートされたアニーリングを含む。加えて、すべてのサンプリング技法のために、もたらされるジャンプの種類を変化させることができる（例えば、ランダムジャンプをランダムジャンプ、バイアスジャンプに（例えば野生型にまたはそれから）、バイアス残基へジャンプ（例えば類似残基にまたはそれから、など））。多重残基位置がカップリングされた場合のジャンプ（2つの残基がいつも一緒に変化し、または組換えのない）、前記の全体のセットが他の配列に変化する場合のジャンプ（例えば組換え）。同様に、すべてのサンプリング技法のために、サンプリングジャンプが許容されるかいなかの許容基準を変化させ、高温で広い配列および低温でローカル最適

40

50

値へ近い狭い配列を可能とすることができる。引用によりここに含めるMetropolis et al., J. Chem Phys v21, pp 1087, 1953参照。

【0090】

加えて、本発明の好ましい方法は、配列のランク整列リストをもたらすということが注意すべきである；すなわち、配列をいくつかの目的基準に基づいてランク化する。しかし、ここに概説のように、非整列配列のセットを、例えばそれらをランク化することなく配列を列挙する、直接的に確率表を生成することによって（例えばSCMF分析または配列アライメント技法を使用して）創出することが可能である。この概説のサンプリング技法を、いずれかの状況で使用できる。

【0091】

好ましい実施態様では、Boltzmanサンプリングをなし得る。当業者に認識されるように、Boltzmanサンプリングのための温度基準を変化させ、高温で広いサーチおよび低温でローカル最適値に近い狭いサーチをもたらすことができる（例えばMetropolis et al., J. Chem. Phys. 21:1087,1953参照）。

【0092】

好ましい実施態様では、サンプリング技法は、遺伝的アルゴリズム、例えばHolland(Adaptation in Natural and Artificial Systems, 1975, Ann Arbor, U. Michigan Press)によって記載されるようなものを利用する。遺伝的アルゴリズム分析は一般に、生成した配列を取り、そして「遺伝子シャッフリング」と類似の態様で、核酸組換え事象と同様に、それらをコンピューター的に組み合わせる。加えて、以下に概説のように、相関多重ジャンプをまたなし得る。そのようなジャンプは、種々の乗り換え位置および1より多い組換え時間で発生でき、そして2または3以上の配列の組換えに関係できる。さらに、欠失または挿入（ランダムまたはバイアス）をなし得る。加えて、以下に概説のように、遺伝的アルゴリズム分析をまた、二次的ライブラリーが生成した後に使用し得る。

【0093】

好ましい実施態様では、サンプリング技法は、シミュレート化アニーリング、例えばKirkpatrick et al.[Science, 220:671-680(1983)]によって記載のように利用する。シミュレートされたアニーリングは、温度を変化することによって、よいまたは悪いジャンプを許容するためのカットオフを変化させる。すなわち、カットオフのストリンジェンシーは、温度を変化させることによって変化させる。これは、配列スペースの新しい領域への、高温で広いサーチを可能とし、低温での狭いサーチに変化して領域を詳細に探索する。

【0094】

加えて、すべて引用によりここに含める米国出願09/9277920、60/352103および60/351937に記載の用に使用し得るコンピューター的方法がある。

任意のタンパク質設計サイクルを、個々に、他の方法と組み合わせて、または組み合わせ方法に関連して使用することができる。

【0095】

好ましい実施態様では、本発明の方法は、標的タンパク質で出発することに関係し、そして実験的方法を使用し、変異体タンパク質を生成する。すなわち、当業者に既知のような核酸組み合わせ技法を使用し、本発明の変異体タンパク質を実験的に生成する。

【0096】

こうして、核酸組換え方法の使用またはタンパク質設計サイクルの履行、または核酸組換え方法およびコンピューター的加工の組み合わせは、変化したコファクター特異性を呈する変異体タンパク質の生成という結果となる。「変異体タンパク質」または「可変タンパク質配列」によってここに、少なくとも1アミノ酸残基で、スキャフォールドタンパク質または標的タンパク質と異なるタンパク質を意味する。

【0097】

好ましい実施態様では、変異体タンパク質のコファクター特異性は変化し、標的タンパク質を比較する。標的タンパク質は限定しないが、チオレドキシシンレダクターゼ、グルタチオンレダクターゼ、および2,5-ジケト-D-グルクロン酸レダクターゼを含む。2つの特異的

10

20

30

40

50

アミノ酸残基は以前、コファクター特異性について報告された (Caugo and Argos, タンパク質s(1997)28,10-28)。グリシンリッチループにすぐに続く第1領域はG-x-G-x-X₁-X₂モチーフを有し、そしてピリジンスクレオチド結合に関係する。本来は、NADPHに特異的なタンパク質では、X₁ およびX₂ はそれぞれ極性残基(Ser/Thr)およびAlaであり、一方NADHに特異的なタンパク質では、X₁ およびX₂ はそれぞれ疎水性残基(Val/Ile) およびGlyである。追加的配列の決定は、しかし、X₁およびX₂についての顕著な配列可変性を実証し、コファクター特異性についてのこの本来のルールを破壊した。

【0098】

第2領域は、E.coli チオレドキシンレダクターゼ中の約175ないしアミノ酸181の領域に一般に対応することが報告されている。NADH依存性細菌性フラボタンパク質レダクターゼCp34およびAhpF (Reynolds et al., Biochemistry (2002) 41, 1990-2001)において、第2モチーフは、それぞれH-Q-F-x-x-x-Q および E-F-A-x-x-x-Kと報告されている。突然変異研究では (Scrutton et al., Nature (1990) 343, 38-43; Mittl et al., タンパク質 Sci. (1994) 3, 1504-1514)、E.coli GRのNADPH特異性は、第2モチーフの、E-M-F-x-x-x-x-P への突然変異によってNADHに切り替えられた(以下の図参照)。

10

【0099】

好ましい実施態様では、変異体チオレドキシンレダクターゼを、コファクター特異性が変化するように作成する。チオレドキシン (TR)は、ほとんどの生物に見出される強力なタンパク質ジスルフィドレダクターゼであり、多くのチオール依存性細胞性還元過程に参加する。細胞性タンパク質の還元をもたらすその能力に加えて、チオレドキシンレダクターゼは抗酸化物質として直接的に作用できる (例えば反応性酸素種を掃討することによって酸化可能基質の酸化を防止することによって) または自己酸化によって細胞中の酸化的ストレスを増加できる (例えば自己酸化を介するスーパーオキシドを生成することによって) ことが認識される。

20

【0100】

チオレドキシンは、低分子量ジチオールタンパク質であって、典型的な有機化合物中のジスルフィド、例えばEllmanの試薬またはこれらが種々のタンパク質中に天然に存在するようなジスルフィドを還元する能力を有する (Holmgren, A. (1981) Trends in Biochemical Science, 6, 26-39)。通常の生理学的条件下で、ジスルフィド結合の還元が続いて、酸化されたチオレドキシンは、コファクターとしてNADPHの助けを借りて、チオレドキシンレダクターゼによって還元される。種のチオレドキシンは典型的に、同じ種のチオレドキシンレダクターゼによってのみ還元される。

30

【0101】

チオレドキシンレダクターゼの活性部位ポケットは、種間で高度に保存された領域を停止、図1Aのアミノ酸アラインメントに示すとおりである。この領域は、E. coliチオレドキシンレダクターゼの残基156と181、またはアラビドプシスチオレドキシンレダクターゼの残基220と245の間のアミノ酸領域に対応する。この高度に保存されたポケットは、コファクター、NADPHの結合のため最大に原因となる。チオレドキシンレダクターゼのアミノ酸配列における種間変異は、CおよびN末端領域すなわちE. coliチオレドキシンレダクターゼの残基1-155と182-C末端およびアラビドプシスチオレドキシンレダクターゼの残基1-219および246-C末端にみえる。

40

【0102】

本発明の変異体チオレドキシンレダクターゼを生成するため使用される標的タンパク質は、限定しないがE. coli、Bacillus subtilis、Mycobacterium leprae、Sarccharomyces、Neurospora crassa、Arabidopsis、Homo sapiens、Methanosarcina acetivorans str. C2A、Ureaplasma parvum、Mycoplasma pulmonis、Rickettsia conorii、Spiroplasma burkhanus、Listeria innocua、Fusobacterium nucleatum、Methanococcus jannaschii、Mycoplasma genitalium、Haemophilus influenzae、Vibrio cholera、Listeria monocytogenes、Helicobacter pylori、Methanopyrus kandleri AV19、Schizosaccharomyces pombe、Chlamydomonas reinhardtii、Streptococcus pyogenes、Plasmodium falciparum、Mycob

50

acterium tuberculosis、Mycoplasma genitalium、Borrelia burgdorferi、Ralstonia solanacearum、Sinorhizobium meliloti、Caulobacter crescentus CB15]、Encephalitozoon cuniculi、Staphylococcus aureus、Clostridium perfringens、Halobacterium sp. NR C-1、Sulfolobus solfataricus、Rickettsia prowazekii、Mesorhizobium loti、Mus musculus、Thermoplasma acidophilum、Sulfolobus tokodaii、Chlamydophila pneumoniae、Mycoplasma pulmonis、Campylobacter jejuni、Chlamydia trachomatis、Aeropyrum pernix、Neisseria meningitidis、Pyrococcus horikoshii、Pyrococcus abyssi、Thermoplasma volcanium、Pyrococcus furiosus、Archaeoglobus fulgidus、Yersinia pestis、Bacillus halodurans、Ureaplasma urealyticum、Methanothermobacter thermautotrophicus、Pyrobaculum aerophilum、Chlamydia muridarum、Treponema pallidum、Streptomyces coelicolor、Brucella melitensis、Agrobacterium tumefaciens、Drosophila melanogaster、Streptococcus pneumoniae、Clostridium acetobutylicum、Xylella fastidiosa、Lactococcus lactis、Thermotoga maritima、Pseudomonas aeruginosa、Salmonella enterica、Nostoc sp、Deinococcus radiodurans、Penicillium chrysogenum、Salmonella typhimurium、Lactobacillus elbrueckii、Clostridium sticklandii、Clostridium litorale、Clostridium acetobutylicum、Thermoplasma volcanium、Rattus norvegicus、Coccidioides immitis、Bos Taurus、Mycobacterium smegmatis、Synechocystis sp、Plasmodium falciparum、Carboxydothemus hydrogenoformans、Sus scrofa Triticum aestivumを含む任意の生物から取得し得る。

10

【0103】

20

好ましい実施態様では、変異体チオレドキシンレダクターゼを生成するために使用する標的タンパク質は、E. coli、Bacillus subtilis、Mycobacterium leprae、Saccharomyces、Neurospora crassa、Arabidopsis、Homo sapiens、米国特許6380372、名称Barley gene for Thioredoxin and NADP-チオレドキシンレダクターゼ、2002年04月30日発行に見出されるオオムギTR; W00198509にアミノ酸配列の配列番号27として、およびそこにそのヌクレオチド配列の配列番号25として見出されるイネTR、Archaeoglobusfulgidus (gil2649006) (trxB)由来のタンパク質配列は、W00198509の配列番号7である熱安定性TR、およびW00198509の配列番号6のMethanococcus jannaschii (gil 1592167) (trxB)由来TRのタンパク質配列から選択される。

【0104】

30

好ましい実施態様では、変異体TRの触媒効率は、コファクターNADPHについて改善される。好ましくは、変異体TRの触媒効率は、NADPHについての野生型と比較して、少なくとも約5%改善される。より好ましくは変異体TRの触媒効率は、NADPHについての野生型と比較して少なくとも約15%改善される。より好ましくは変異体TRの触媒効率は、NADPHについての野生型と比較して少なくとも約15%改善される。より好ましくは変異体TRの触媒効率は、NADPHについての野生型と比較して少なくとも約25%改善される。より好ましくは変異体TRの触媒効率は、NADPHについての野生型と比較して少なくとも約50%改善される。より好ましくは変異体TRの触媒効率は、NADPHについての野生型と比較して少なくとも約100%改善される。より好ましくは変異体TRの触媒効率は、NADPHについての野生型と比較して少なくとも約300%改善される。より好ましくは変異体TRの触媒効率は、NADPHについての野生型と比較して少なくとも約500%改善される。

40

【0105】

好ましい実施態様では、変異体TRタンパク質の触媒効率は、コファクターNADHについて改善される。好ましくは変異体TRの触媒効率は、NADHについての野生型と比較して少なくとも約5%改善される。より好ましくは変異体TRの触媒効率は、NADHについての野生型と比較して少なくとも約15%改善される。より好ましくは変異体TRの触媒効率は、NADHについての野生型と比較して少なくとも約25%改善される。より好ましくは変異体TRの触媒効率は、NADHについての野生型と比較して少なくとも約50%改善される。より好ましくは変異体TRの触媒効率は、NADHについての野生型と比較して少なくとも約100%改善される。より好ましくは変異体TRの触媒効率は、NADHについての野生型と比較して少なくとも約300%改善

50

される。より好ましくは変異体TRの触媒効率は、NADHについての野生型と比較して少なくとも約500%改善される。より好ましくは変異体TRの触媒効率は、NADHについての野生型と比較して少なくとも約1000%改善される。より好ましくは変異体TRの触媒効率は、NADHについての野生型と比較して少なくとも約1300%改善される。より好ましくは変異体TRの触媒効率は、NADHについての野生型と比較して少なくとも約3000%改善される。

【0106】

好ましい実施態様では、変異体チオレドキシンレダクターゼのコファクター特異性を変化させ、その結果NADHを使用する活性増加がある。好ましくは、変異体チオレドキシンレダクターゼ(TR)は、NADHを使用するもとのNADPH依存性活性の少なくとも50%を有する。より好ましくは、変異体チオレドキシンレダクターゼ(TR)は、NADHを使用するもとのNADPH依存性活性の少なくとも75%を有する。より好ましくは、変異体チオレドキシンレダクターゼ(TR)は、NADHを使用するもとのNADPH依存性活性の少なくとも85%を有する。より好ましくは、変異体チオレドキシンレダクターゼ(TR)は、NADHを使用するもとのNADPH依存性活性の少なくとも95%を有する。より好ましくは、変異体チオレドキシンレダクターゼ(TR)は、NADHを使用するもとのNADPH依存性活性の少なくとも100%を有する。

10

【0107】

好ましい実施態様では、変異体チオレドキシンレダクターゼのコファクター特異性は変化し、その結果NADPHからNADHへのコファクター切り替えがある。他の言葉では、これらの変異体は、NADH依存性活性の増加および実質的に同時的な、NADPH依存性活性の減少を有する。好ましくは、変異体チオレドキシンレダクターゼ (TR)は、NADHを使用するもとのNADPH依存性活性の少なくとも50%を有する。より好ましくは変異体チオレドキシンレダクターゼは、NADHを使用するもとのNADPH依存性活性の少なくとも75%を有する。より好ましくは変異体チオレドキシンレダクターゼは、NADHを使用するもとのNADPH依存性活性の少なくとも85%を有する。より好ましくは変異体チオレドキシンレダクターゼは、NADHを使用するもとのNADPH依存性活性の少なくとも95%を有する。より好ましくは変異体チオレドキシンレダクターゼは、NADHを使用するもとのNADPH依存性活性の少なくとも100%を有する。より好ましくは変異体チオレドキシンレダクターゼは、NADHを使用するもとのNADPH依存性活性の少なくとも75%を有する。

20

【0108】

好ましくは変異体チオレドキシンレダクターゼ (TR)は、もとのNADPH依存性活性の約0.5%より小さい活性を有する。より好ましくは、TRは、もとのNADPH依存性活性の約5%より小さい活性を有する。より好ましくは、TRは、もとのNADPH依存性活性の約20%より小さい活性を有する。より好ましくは、TRは、もとのNADPH依存性活性の約25%より小さい活性を有する。より好ましくは、TRは、もとのNADPH依存性活性の約30%より小さい活性を有する。より好ましくは、TRは、もとのNADPH依存性活性の約50%より小さい活性を有する。より好ましくは、TRは、もとのNADPH依存性活性の約75%より小さい活性を有する。より好ましくは、TRは、もとのNADPH依存性活性の約95%より小さい活性を有する。

30

【0109】

さらなる実施態様では、変異体TRタンパク質の触媒効率は両方のコファクター、NADHおよびNADPH一緒について、改善される。好ましくはTR変異体の触媒効率は、2つのコファクターのいずれかについての野生型と比較して、少なくとも約5%改善される。より好ましくはTR変異体の触媒効率は、2つのコファクターのいずれかについての野生型と比較して、少なくとも約50%改善される。より好ましくはTR変異体の触媒効率は、2つのコファクターのいずれかについての野生型と比較して、少なくとも約100%改善される。より好ましくはTR変異体の触媒効率は、2つのコファクターのいずれかについての野生型と比較して、少なくとも約300%改善される。より好ましくはTR変異体の触媒効率は、2つのコファクターのいずれかについての野生型と比較して、少なくとも約1000%改善される。より好ましくはTR変異体の触媒効率は、2つのコファクターのいずれかについての野生型と比較して、少なくとも約2000%改善される。

40

【0110】

50

好ましい実施態様では、変異体チオレドキシンレダクターゼ (TR) の NADPH 結合親和性は影響、減少または増強されなくてもよい。例えば、いくつかの実施態様では変異体 TR は NADH についてよりも NADPH について 100 倍より親和性であり得、一方他の実施態様では、変異体 TR は、NADH についてよりも NADPH について 50 倍より親和性を示す。または変異体 TR は NADH についてよりも NADPH について 25 倍より大きい親和性を示し得る。

好ましい実施態様では、変異体 TR タンパク質の、そのコグネートチオレドキシンを還元する能力は、実質的に影響を受けない。

【0111】

好ましい実施態様では、変異体 TR タンパク質の基質特異性は、影響を受け、その結果 TR タンパク質は、他の種からのチオレドキシンタンパク質に作用し得る。

10

幾つかの実施態様では、タンパク質付加サイクルによって付加された潜在的グリコシル化部位を、第2タンパク質設計サイクルを使用することによって活性に影響することなく除去し得る。

【0112】

好ましい実施態様では、変異体 TR タンパク質は、野生型 TR タンパク質と比較して、コファクター特異性に関するアミノ酸における1ないし3アミノ酸置換を有する。他の実施態様では、変異体 TR タンパク質は、他の部分に追加的アミノ酸置換を有する。こうして、変異体 TR タンパク質は他の位置に、少なくとも2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40つの異なる残基を有し得る。当業者に認識されるように、アミノ酸置換を有し得る追加的位置の数は、変異体を生成するために使用される野生型 TR タンパク質に依存する。こうして、幾つかの例では、50までの異なる位置は、アミノ酸置換を有し得る。

20

【0113】

好ましい実施態様では、変異体 TR タンパク質は、アラビドプシス NTR タンパク質 (Genbank 受託番号 Q39242) の位置 190、191、および 195、および E.coli TR タンパク質 (Genbank 受託番号 P09625) の 175、Bacillus subtilis TR タンパク質 (Genbank 受託番号 P80880) の位置 155、156、および 174、Mycobacterium leprae TR タンパク質 (Genbank 受託番号 P46843) の位置 163、164、および 182、Sacchromyces TR タンパク質 (Genbank 受託番号 P29509 および P38816) の残基 164、165、および 183、Neurospora crassa TR タンパク質 (Genbank 受託番号 P51978) の位置 163、164、および 182、アラビドプシス TR タンパク質 (Genbank 受託番号 Q39243) の残基 170、171、189 および ヒト TR タンパク質 (Genbank 受託番号 Q16881) の残基 217、218 および 249 に対応する位置 A4、A5 および A6 から選択されるアミノ酸置換を含む。

30

【0114】

好ましい実施態様では、変異体 TR タンパク質は、RA4W、RA5L、RA5M、RA5I、RA5F、RA5V、RA5Y、RA5A、RA5S、RA5C、RA5T、RA6T、RA6S、RA6Q、RA6G、および RA6N、RA6D、RA6M、および RA6E からなる置換の群より選択されるアミノ酸置換を含む。

【0115】

好ましい実施態様では、変異体 TR タンパク質は、アミノ酸置換 RA4W および RA6T を含む。
好ましい実施態様では、変異体 TR タンパク質は、アミノ酸置換 RA4W、RA5L、および RA6S を含む。

40

【0116】

好ましい実施態様では、変異体 TR タンパク質は、アミノ酸置換 RA5Y および RA6N を含む。
好ましい実施態様では、変異体 TR タンパク質は、アミノ酸置換 RA4W、RA5F、および RA6Q を含む。
好ましい実施態様では、変異体 TR タンパク質は、アミノ酸置換 RA4W、RA5L、および RA6T を含む。

【0117】

好ましい実施態様では、変異体 TR タンパク質は、アミノ酸置換 RA4W および RA6S を含む。

50

好ましい実施態様では、変異体TRタンパク質は、アミノ酸置換RA5YおよびRA6Nを含む。

好ましい実施態様では、変異体TRタンパク質は、アミノ酸置換RA5FおよびRA6Nを含む。

好ましい実施態様では、変異体TRタンパク質は、アミノ酸置換RA4WおよびRA6Tを含む。

【0118】

好ましい実施態様では、変異体TRタンパク質は、アミノ酸置換RA4W、RA5LおよびRA6Sを含む。

好ましい実施態様では、変異体TRタンパク質は、アミノ酸置換RA4W、RA5M、およびRA6Sを含む。

好ましい実施態様では、変異体TRタンパク質は、アミノ酸置換RA4W、RA5I、およびRA6Sを含む。

好ましい実施態様では、変異体TRタンパク質は、アミノ酸置換RA4W、RA5F、およびRA6Qを含む。

10

【0119】

好ましい実施態様では、変異体TRタンパク質は、アミノ酸置換RA4W、およびRA5Vを含む。

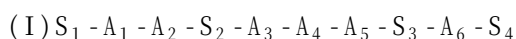
好ましい実施態様では、変異体TRタンパク質は、アミノ酸置換RA4W、RA5M、およびRA6Gを含む。

好ましい実施態様では、変異体TRタンパク質は、アミノ酸置換RA4W、RA5V、およびRA6Gを含む。

【0120】

好ましい実施態様では、変異体タンパク質は、式Iのポリペプチド分子である。

20



ここで、

a) S_1 は配列番号1、配列番号2、配列番号3、配列番号4、配列番号5、配列番号6、および配列番号7、またはそれに実質的類似性を有する配列からなる群より選択されるポリペプチドを含む；

b) S_2 は配列番号8、配列番号9、配列番号10、配列番号11、配列番号12、配列番号13、および配列番号14、またはそれに実質的類似性を有する配列からなる群より選択されるポリペプチドを含む；

【0121】

c) S_3 は配列番号15、配列番号16、配列番号17、配列番号18、配列番号19、配列番号20、および配列番号21、またはそれに実質的類似性を有する配列からなる群より選択されるポリペプチドを含む；

30

d) S_4 は配列番号22、配列番号23、配列番号24、配列番号25、配列番号26、配列番号27、および配列番号28、またはそれに実質的類似性を有する配列からなる群より選択されるポリペプチド配列を含む；

e) A_1 は、セリン、バリン、グリシン、アラニン、ロイシン、イソロイシン、メチオニン、フェニルアラニン、およびトリプトファンからなる群より選択されるアミノ酸部分である；

【0122】

f) A_2 はアラニン、グリシン、バリン、ロイシン、イソロイシン、メチオニン、フェニルアラニン、およびトリプトファンからなる群より選択されるアミノ酸部分である；

40

g) A_3 はヒスチジン、アスパラギン酸、グルタミン酸、アルギニン、ロイシン、セリン、トレオニン、システイン、アスパラギン、グルタミン、およびチロシンからなる群より選択されるアミノ酸部分である；

h) A_4 はアルギニン、アラニン、グリシン、バリン、ロイシン、イソロイシン、メチオニン、フェニルアラニン、およびトリプトファンからなる群より選択されるアミノ酸部分である；

i) A_5 はアルギニン、アスパラギン、グルタミン、アスパラギン酸、グルタミン酸、システイン、セリン、トレオニン、およびリジンからなる群より選択されるアミノ酸部分である；

50

【0123】

j) A_6 はアルギニン、グルタミン酸、アスパラギン、グルタミン、アスパラギン酸、システイン、セリン、トレオニン、およびリジンからなる群より選択されるアミノ酸部分である；

【0124】

ただし、少なくとも

A_1 は、セリンではない；

A_2 は、アラニンではない；

A_3 は、ヒスチジンではない；

A_4 は、アルギニンではない；

A_5 は、アルギニンではない；または

A_6 は、アルギニンではない。

10

【0125】

前記式Iでは、配列 A_1 - A_2 - S_2 - A_3 - A_4 - A_5 - S_3 - A_6 は種々の種から取得されるチオレドキシンレダクターゼタンパク質の配列の高度に保存されたポケットに対応する。 A_1 は、*E. coli*チオレドキシンレダクターゼ配列の残基156、*Bacillus subtilis*チオレドキシンレダクターゼ配列の残基155、*Mycobacterium leprae*チオレドキシンレダクターゼ配列の残基163、*Sarccharomyces*チオレドキシンレダクターゼ配列の残基164、*Neurospora crassa*チオレドキシンレダクターゼ配列の残基163、アラビドプシスチオレドキシンレダクターゼ配列の残基170、およびヒトチオレドキシンレダクターゼ配列の残基217に対応する。野生型タンパク質では、この残基は*E. coli*およびヒトではトレオニン、および他の前列举生物ではセリンである。

20

【0126】

A_2 は、*E. coli*チオレドキシンレダクターゼ配列の残基157、*Bacillus subtilis*チオレドキシンレダクターゼ配列の残基156、*Mycobacterium leprae*チオレドキシンレダクターゼ配列の残基164、*Sarccharomyces*チオレドキシンレダクターゼ配列の残基165、*Neurospora crassa*チオレドキシンレダクターゼ配列の残基164、アラビドプシスチオレドキシンレダクターゼ配列の残基171、ヒトチオレドキシンレダクターゼ配列の残基218に対応する。野生型タンパク質では、この残基は、ヒトではバリンおよび他の列举種全てではアラニンである。

30

【0127】

A_3 は、*E. coli*チオレドキシンレダクターゼ配列の残基175、*Bacillus subtilis*チオレドキシンレダクターゼ配列の残基174、*Mycobacterium leprae*チオレドキシンレダクターゼ配列の残基182、*Sarccharomyces*チオレドキシンレダクターゼ配列の残基183、*Neurospora crassa*チオレドキシンレダクターゼ配列の残基182、アラビドプシスチオレドキシンレダクターゼ配列の残基189、ヒトチオレドキシンレダクターゼ配列の残基249に対応する。野生型タンパク質では、この残基は、ヒトではアルギニン、*Sarccharomyces*および*Neurospora crassa*ではバリン、および他のすべての列举生物ではヒスチジンである。

【0128】

A_4 は、*E. coli*チオレドキシンレダクターゼ配列の残基残基176、*Bacillus subtilis*チオレドキシンレダクターゼ配列の残基175、*Mycobacterium leprae*チオレドキシンレダクターゼ配列の残基183、*Sarccharomyces*チオレドキシンレダクターゼ配列の残基184、*Neurospora crassa*チオレドキシンレダクターゼ配列の残基183、アラビドプシスチオレドキシンレダクターゼ配列の残基190、ヒトチオレドキシンレダクターゼ配列の残基250に対応する。野生型タンパク質では、この残基はヒトではグルタミンおよび他のすべての列举生物ではアルギニンである。

40

【0129】

A_5 は、*E. coli*チオレドキシンレダクターゼ配列の残基177、*Bacillus subtilis*チオレドキシンレダクターゼ配列の残基176、*Mycobacterium leprae*チオレドキシンレダクターゼ配列の残基184、*Sarccharomyces*チオレドキシンレダクターゼ配列の残基185、*Neurospo*

50

ra crassaチオレドキシンレダクターゼ配列の残基184、アラビドプシスチオレドキシンレダクターゼ配列の残基191、ヒトチオレドキシンレダクターゼ配列の残基251に対応する。野生型タンパク質では、この残基はSarccharomycesおよびNeurospora crassaではリジン、ヒトではフェニルアラニン、および他のすべての列挙生物ではアルギニンである。

【0130】

A₆は、E. coli チオレドキシンレダクターゼ配列の残基、Bacillus subtilisチオレドキシンレダクターゼ配列の残基180、Mycobacterium lepraeチオレドキシンレダクターゼ配列の残基188、Sarccharomycesチオレドキシンレダクターゼ配列の残基189、Neurospora crassaチオレドキシンレダクターゼ配列の残基188、アラビドプシスチオレドキシンレダクターゼ配列の残基195、ヒトチオレドキシンレダクターゼ配列の残基255に対応する。野生型タンパク質では、この残基は、ヒトではリジンおよび他のすべての列挙生物ではアルギニンである。 10

【0131】

前記種間で、S₂およびS₃に対応するアミノ酸配列の部分はまた、高度に保存されていることが観察された。S₂は配列番号8、配列番号9、配列番号10、配列番号11、配列番号12、配列番号13、および配列番号14からなる群より選択されるポリペプチド配列を含む。S₃は配列番号15、配列番号16、配列番号17、配列番号18、配列番号19、配列番号20、および配列番号21からなる群より選択されるポリペプチド配列を含む(図2)。

【0132】

したがって、本発明の実施態様は、式Iのポリペプチドに関し、ここでS₁は配列番号1、配列番号2、配列番号3、配列番号4、配列番号5、配列番号6、および配列番号7からなる群より選択される配列を有するポリペプチドからなる。 20

【0133】

ある種の実施態様では、S₂は配列番号8、配列番号9、配列番号10、配列番号11、配列番号12、配列番号13、および配列番号14からなる群より選択されるポリペプチド配列からなる。一方、S₃は配列番号15、配列番号16、配列番号17、配列番号18、配列番号19、配列番号20、および配列番号21からなる群より選択されるポリペプチド配列からなる。本発明の他の実施態様は、S₄は、配列番号22、配列番号23、配列番号24、配列番号25、配列番号26、配列番号27、および配列番号28からなる群より選択される配列を有するポリペプチド配列からなる。 30

【0134】

1実施態様では、式Iのポリペプチドでは、S₁は配列番号1に示すポリペプチド配列であり、S₂は、配列番号8に示すポリペプチド配列であり、S₃は、配列番号15に示すポリペプチド配列であり、およびS₄は、配列番号22に示すポリペプチド配列である。これは、E.coliから取得される、チオレドキシンレダクターゼタンパク質、またはその突然変異体、に対応する。

【0135】

べつの実施態様では、式Iのポリペプチドでは、S₁は、配列番号2に示すポリペプチド配列であり、S₂は、配列番号9に示すポリペプチド配列であり、S₃は、配列番号16に示すポリペプチド配列であり、およびS₄は、配列番号23に示すポリペプチド配列である。これは、Bacillus subtilisから取得されるチオレドキシンレダクターゼタンパク質、またはその突然変異体に対応する。 40

【0136】

なおさらなる実施態様では、式Iのポリペプチドで、S₁は、に示すポリペプチド配列であり配列番号3、S₂はthe は、に示すポリペプチド配列であり配列番号10、S₃はthe は、に示すポリペプチド配列であり配列番号17、およびS₄は、配列番号24に示すポリペプチド配列である。これは、Mycobacterium lepraeから取得されるチオレドキシンレダクターゼタンパク質、またはその突然変異体に対応する。

【0137】

本発明の別の実施態様では、式Iのポリペプチドに関し、ここでS₁は、配列番号4に示すポ 50

リペプチド配列であり、 S_2 は、配列番号11に示すポリペプチド配列であり、 S_3 は、配列番号18に示すポリペプチド配列であり、および S_4 は、配列番号25に示すポリペプチド配列であり。これは、*Sarccharomyces*から取得されるチオレドキシンレダクターゼタンパク質、またはその突然変異体に対応する。

【0138】

別の実施態様では、式Iのポリペプチドで、 S_1 は、配列番号5に示すポリペプチド配列であり、 S_2 は、配列番号12に示すポリペプチド配列であり、 S_3 は、配列番号19に示すポリペプチド配列であり、および S_4 は、配列番号26に示すポリペプチド配列である。これは、*Neurospora crassa*から取得されるチオレドキシンレダクターゼタンパク質、またはその突然変異体に対応する。

10

【0139】

1実施態様では、式Iのポリペプチドでは、 S_1 は、配列番号6に示すポリペプチド配列であり、 S_2 は、配列番号13に示すポリペプチド配列であり、 S_3 は、配列番号20に示すポリペプチド配列であり、および S_4 は、配列番号27に示すポリペプチド配列である。これは、アラビドプシスから取得されるチオレドキシンレダクターゼタンパク質、またはその突然変異体に対応する。

【0140】

本発明はまた、式Iのポリペプチドに関し、ここで S_1 は、配列番号7に示すポリペプチド配列であり、 S_2 は、配列番号14に示すポリペプチド配列であり、 S_3 は、配列番号21に示すポリペプチド配列であり、および S_4 は、配列番号28に示すポリペプチド配列である。これは、ヒトから取得されるチオレドキシンレダクターゼタンパク質、またはその突然変異体に対応する。

20

【0141】

本発明は天然存在チオレドキシンレダクターゼタンパク質のある種の突然変異体を包含する。これらの突然変異体は、以下のものを含みここで A_1 はバリン、アラニン、およびロイシンからなる群より選択されるアミノ酸部分である； A_2 はグリシン、バリン、およびロイシンからなる群より選択されるアミノ酸部分である； A_3 はアスパラギン酸、グルタミン酸、アスパラギン、およびグルタミンからなる群より選択されるアミノ酸部分である； A_4 はアラニン、グリシン、バリン、ロイシン、イソロイシン、およびメチオニンからなる群より選択されるアミノ酸部分である； A_5 はアスパラギン、グルタミン、アスパラギン酸、およびグルタミン酸からなる群より選択されるアミノ酸部分である； A_6 はグルタミン酸、グルタミン、アスパラギン酸、およびアスパラギンからなる群より選択されるアミノ酸部分である。

30

【0142】

本発明は前記突然変異体のいずれかの1または2以上を有し得ることが理解される。ある種の実施態様では、 A_1 がバリンであり、一方他には A_2 はグリシンであり、そして他には A_3 はアスパラギン酸である；そして他に A_4 はアラニンであり、一方他に A_5 はアスパラギンであり、そして他に A_6 はグルタミン酸である。いくつかの好ましい実施態様では、これら特定のアミノ酸残基の2または3以上が特定の位置に存在する。

【0143】

好ましい実施態様では、本発明の変異体タンパク質は、第2タンパク質に融合し得る。例えば、式Iのポリペプチドおよび第2ポリペプチドを含む融合タンパク質を作成し得る。第2ポリペプチドは、野生型TRタンパク質、野生型チオレドキシン、またはタンパク質設計サイクルによって生成される変異体タンパク質であり得る。代替的に、タンパク質設計サイクルによって生成される変異体タンパク質および第2ポリペプチドを含む融合タンパク質を融合し得る。第2ポリペプチドは野生型TRタンパク質、野生型チオレドキシンまたは式Inoポリペプチドであり得る。そのような融合はリンカーを経由し得る。

40

【0144】

ここで「リンカー」、「リンカー配列」、「スペーサー」またはその文法的均等物によって、2分子を接続しおよび好ましい構成で2分子を配置するようにしばしば供する1分子

50

または分子の群（例えばモノマーまたはポリマー）を意味する。この実施態様の1側面では、リンカーはペプチド結合である。2ポリペプチド鎖が種々のパラメーター、例えば2ポリペプチド鎖（例えばこれらが天然にダイマーを形成し、またはしない）、もし3次元構造決定からしられるならば接続すべきであるNおよびC末端の間の距離および/またはタンパク質溶解および酸化に向かう安定性に依存して接続すべき特異の場合について、好適なリンカーを選択すること。さらに、リンカーはフレキシビリティを提供する網の酸残基を含み得る。こうして、リンカーペプチドは、以下のアミノ酸残基:Gly、Ser、Ala、またはThrを卓越して含み得る。

【0145】

リンカーペプチドは、2モノマーを、これらが互いについて正確なコンホメーションを確保するようなやり方で連結するに十分である長さを有すべきであり、その結果これらは所定のレセプターのアンタゴニストとしての望まれる活性を保持する。この目的のため好適な長さは、数なくとも1および30アミノ酸残基を超えない。好ましくは、該リンカーは、約1ないし30アミノ酸長であり、1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19および20アミノ酸長のリンカーが好ましい。また引用によりここに含めるW001/25277参照。

10

【0146】

加えて、リンカーペプチドに含まれるためのアミノ酸残基は、ポリペプチドの活性を顕著には妨害しない特性を呈すべきである。こうして、全体のリンカーペプチドは、ポリペプチドの活性と両立しない、またはフォールディングに干渉しない、またはレセプターモノマードメインの結合を重度に抵抗する1または2以上のモノマーのアミノ酸残基の結合または他の相互作用を形成する変化を呈すべきでない。

20

【0147】

有用なリンカーはグリシン-セリンポリマー（例えば $(GS)_n$ 、 $(GS_2GS)_n$ 、 $(GGGS)_n$ および $(GGGS)_n$ を含みここでnは少なくとも1の整数である）、グリシン-アラニンポリマー、アラニン-セリンポリマー、および他のフレキシブルリンカーたとえばシェーカーカリウムチャンネルのための繋ぎ、および種々の他のフレキシブルリンカーを含み、当業者に認識される通りである。グリシン-セリンポリマーは、これらの両方のアミノ酸が比較的に剛構造であれう化ら、好ましくあり得、したがって、コンポーネント間の中世繋ぎとして供し得る。第2に、セリンは親水性であり、したがって、グロブラーグリシン鎖であり得るものを可溶化し得る。第3に、類似の鎖は、組換えタンパク質のサブユニット、例えば単一鎖抗体を結合するのに有効であると示された。

30

【0148】

好適なリンカーはまた、既知3次元構造のデータベースを、2ポリペプチド鎖間のギャップを架橋できる天然存在モチーフについてスクリーニングすることによって同定し得る。好適なリンカーを取得する他のやり方は、単純リンカー例えば $(Gly_4Ser)_n$ をランダム突然変異誘発によって最適化することによる。

【0149】

好ましい実施態様では、リンカーは、約5と約50アミノ酸の間、または約10と40アミノ酸の間または約15と25アミノ酸の間を有するポリペプチド配列を含み得る。好ましい実施態様では、リンカーは約22アミノ酸である。

40

【0150】

好ましい実施態様では、本発明の変異体タンパク質は、第3のポリペプチドに融合され、そしてまた、そのような融合はリンカーを介し得る。融合ポリペプチドの間のリンカーは、式Iのポリペプチドを含み、そして第3ポリペプチドは、約5と約100kDaの間の分子量、または約20と約70kDaの間の分子量、またはなお約25と約45の間の分子量を有し得る。好ましい実施態様では、リンカーは、約30と約40kDaの間の分子量を有する。ある種の実施態様では、このリンカーは、負に荷電した、アミノ酸残基、例えばグルタメートおよびアスパルテートを含む。

【0151】

50

ある種の実施態様では、第3ポリペプチドはオレオシンである。

こうして、本発明の1実施態様は、式Iのポリペプチドに関し、これはそのC末端の第二ポリペプチドに、多分リンカーを介して、融合され、そしてまた、そのN末端で第3ポリペプチドに融合され、また多分さらなるリンカーを介する。本発明のさらなる実施態様は、式IIの1連の融合ポリペプチドに関する。

(II)オレオシン-リンカー1-チオレドキシシンレダクターゼ-リンカー2-チオレドキシシン
ここで、「リンカー1」は、前記のように、式Iのポリペプチドと、第3ポリペプチドの間のリンカーをいい、そして「リンカー2」は前記のように、式Iのポリペプチドと第2ポリペプチドの間のリンカーをいう。同様に、本発明の幾つかの実施態様は、そのN末端、C末端または両方で他のタンパク質に融合されているかいか、式Iのポリペプチドを含む任意の他の融合タンパク質を含む。具体的には、本発明は、式IIまたは式Iのポリペプチドとへの2ポリペプチドの任意の他の融合を企図し、ここで該コンポーネントは、任意の順序で存在する。

10

【0152】

好ましい実施態様では、変異体TRタンパク質の、NADPHおよびANDHについての結合親和性は、決定される。好適なアッセイは限定しないが、例えば、動力学および平行結合定数を比較する定量的比較を含む。動力学結合率 (K_{on}) および解離率 (K_{off})、および平衡結合定数 (K_d) は、文献の標準的手法 [Pearce et al., Biochemistry 38:81-89 (1999)] にしたがって BIAcore 装置における表面プラスモン共鳴を使用して、決定できる。

【0153】

20

好ましい実施態様では、変異体TRタンパク質の宿主動物の抗原性プロファイルは、宿主TRの抗原性プロファイルに類似そして好ましくは同一であり、すなわち、変異体TRタンパク質は有意に、免疫応答への宿主生物（例えば患者）を刺激しない；すなわち、任意の免疫応答は、臨床的に関連せず、そして抗体によるタンパク質のアレルギー応答または中和がない。すなわち好ましい実施態様では、変異体TRタンパク質は、追加的かつTRと異なるエピトープを含まない。「エピトープ」または「決定」によって、ここに、抗体を生成および/または結合するタンパク質の部分を意味する。こうして、ほとんどの例では、抗体の顕著な量は、変異体TRタンパク質を生成しない。一般に、これは表面残基を顕著には変化しないことによって、または任意のアミノ酸残基をグリコシル化されることのできる表面上に追加することによって達成し、新規グリコシル化が免疫応答という結果となる。

30

【0154】

本発明の変異体TRタンパク質および核酸は、天然存在TRと区別可能である。「天然存在」または「野生型」または文法的均等物によってここに、天然に見出されるアミノ酸配列またはヌクレオチド配列を意味し、そしてアレル変異を含む；すなわち、通常意図的には修飾されていないアミノ酸配列またはヌクレオチド配列。よって、「非天然存在」または「合成」または文法的均等物によってここに、天然に見出されないアミノ酸またはヌクレオチド配列を意味する；すなわち通常は意図的には修飾されていないアミノ酸配列またはヌクレオチド配列。ひとたび組換え体核酸が作成されそして宿主細胞または生物に再導入されると、それは非組換え的に、すなわちインビトロ操作でなくむしろ宿主細胞の細胞性機構を使用して複製し、しかしそのような核酸はひとたび組換え的に生産されると、逐次的に非組換え的に複製されるが、本発明の目的のためになお組換え体と考えられる。天然存在TRタンパク質の代表的アミノ酸配列は図21に示す。特記しないかぎり、変異体TRタンパク質および変異体TRタンパク質のすべての位置番号は、これらの配列に基づく、すなわち、当業者に認識されるように、TRタンパク質および変異体TRタンパク質のアラインメントは、標準的プログラムを使用して、なし得、以下に概説の通りであり、2タンパク質の間の「均等」位置の同定による。

40

【0155】

こうして、好ましい実施態様では、変異体TRタンパク質は少なくとも1-5%の残基によって、野生型TR配列（図21）と異なるアミノ酸配列を有する。すなわち、本発明の変異体タンパク質は、野生型TRアミノ酸配列と約97-99%より小さい同一性を有する。よって、もし

50

アミノ酸配列へのタンパク質配列の全体相同性が野生型TRタンパク質の、好ましくは約99%より小さい、より好ましくは約98%より小さい、なおより好ましくは約97%より小さい、そしてより好ましくは95%より小さいならば、タンパク質は「変異体TRタンパク質」である。いくつかの実施態様では、相同性は、約75-80%と同じくらい低い。別にいえば、変異体タンパク質は、野生型TR配列（すなわち図21）と異なる少なくとも約1残基を有し、少なくとも約2、3、4、5、50までの異なる残基である。好ましくは変異体TRタンパク質は、1ないし3この異なる残基を有する。より好ましくは、変異体TRタンパク質は、3ないし5この異なる残基を有する。好ましくは、変異体TRタンパク質は、5ないし10この異なる残基を有する。好ましくは変異体TRタンパク質は、10ないし15個異なる残基を有する。好ましくは変異体TRタンパク質は、15ないし25個異なる残基を有する。好ましくは変異体TRタンパク質は、25ないし35個異なる残基を有する。 10

【0156】

この文脈では、相同性は、配列類似性または同一性を意味し、同一性が好ましい。当業界で知られるように、ある数の異なるプログラムを使用し、タンパク質（または以下議論のように、核酸）が既知配列と配列同一性または類似性を有するか同定できる。配列同一性および/または類似性を、当業界既知標準技法を使用して決定する。限定しないがSmith & Waterman, Adv. Appl. Math., 2:482 (1981)のローカル配列同一性アルゴリズム、Needleman & Wunsch, J. Mol. Biol., 48:443 (1970)の配列同一性アラインメントアルゴリズムにより、Pearson & Lipman, Proc. Natl. Acad. Sci. U.S.A., 85:2444 (1988)の類似性方法についてのサーチにより、これらのアルゴリズムのコンピューター化手段により (Wisconsin Genetics Software Package, Genetics Computer Group, 575 Science Drive, Madison, WIにおけるGAP, BESTFIT, FASTA, およびTFASTA)、Devereux et al., Nucl. Acid Res., 12:387-395 (1984)によって記載されるBest Fit配列プログラム、好ましくはデフォルトセッティングを使用し、または目視によるものを含む。好ましくはパーセント同一性は、以下のパラメーターに基づいてFastDBによって計算する：1のミスマッチペナルティ；1のギャップペナルティ；0.33のギャップサイズペナルティ；および30の結合ペナルティ 「Current Methods in Sequence Comparison and Analysis,」 Macromolecular Sequencing and Synthesis, Selected Methods and Applications, pp 127-149 (1988), Alan R. Liss, Inc. 20

【0157】

有用なアルゴリズムの別の例は、PILEUPである。PILEUPは、進行性、対をなすアラインメントを使用して、関連配列の群から多重配列を創出する。それはまた、アラインメントを創出するために使用されるクラスター化関連性を示すツリーをプロットできる。PILEUPは、Feng & Doolittle, J. Mol. Evol. 35:351-360 (1987)の進行性アラインメント方法；の単純化を使用する；該方法はHiggins & Sharp CABIOS 5:151-153 (1989)によって記載されたものと類似する。有用なPILEUPパラメーターは、3.00のデフォルトギャップ重量、0.10のデフォルトギャップ長重量、および秤量された末端ギャップを含む。 30

【0158】

有用なアルゴリズムのさらなる例は、BLASTアルゴリズムであってAltschul et al., J. Mol. Biol. 215, 403-410, (1990); Altschul et al., Nucleic Acids Res. 25:3389-3402 (1997); およびKarlin et al., Proc. Natl. Acad. Sci. U.S.A. 90:5873-5787 (1993)に記載される。特に有用なBLASTプログラムは、WU-BLAST-2プログラムであってAltschul et al., Methods in Enzymology, 266:460-480 (1996); [http://blast.wustl.edu/blast/ README.html](http://blast.wustl.edu/blast/README.html) から取得される。WU-BLAST-2は、いくつかのサーチパラメーターを使用し、ほとんどのそれはデフォルト値にセットされる。調節可能パラメーターは以下の値でセットされる：オーバーアップスパン=1, オーバーラップフラクション = 0.125, ワードスレシールド (T) = 11. HSP SおよびHSP S2パラメーターは、力学的値であり、そして特定配列の組成および所望の配列がサーチされる特定のデータベースの組成にそれ自体依存するプログラムによって確立される；しかし、値は、感度を増加するために調節し得る。追加的な有用なアルゴリズムは、ギャップ化BLASTであってAltschul et al., Nucl. Acids 40 50

Res., 25:3389-3402によって報告された通りである。ギャップ化BLASTはBLOSUM-62 置換スコア;9にセットされたスレッシュホールドTパラメーター;非ギャップ化伸長をトリガーするための2-ヒット方法; kのギャップ長を負荷するあるコストの $10+k$; X_u は16にセットし, およびデータベースサーチ段階のために X_g は40にセットし、そしてアルゴリズムの出力段階のために67にセットする。ギャップ化アラインメントは、-22ビットに対応するスコアによってトリガーされる。

【0159】

A%アミノ酸配列同一性値は、アラインメントされた領域中の「より長い」配列の残基の総数によって割ったマッチする同一残基の数によって決定される。「より長い」配列は、アラインメントされた領域中のもっとも実際の残基を有するものである（アラインメントスコアを最大化するためにWU-Blast-2によって導入されたギャップは無視する）。

10

【0160】

類似の態様では、ここに同定したポリペプチドのコード配列についての「%核酸配列同一性」は、細胞周期タンパク質のコード配列中のヌクレオチド残基の同一である候補配列中のヌクレオチド残基のパーセンテージとして定義される。好ましい方法は、デフォルトパラメーターに対するWU-BLAST-2セットのBLASTNモジュールを利用し、オーバーラップスパンおよびオーバーラップフラクションセットはそれぞれ1および0.125である。

【0161】

このアラインメントは、アラインメントしようとする配列におけるギャップの導入を含むことができる。さらに、野生型TR配列より多いまたは少ないアミノ酸を含む配列については（即ち、図2、図16Nを参照されたい）、或る態様において、アミノ酸総数に対する一致アミノ酸の数に基づき、配列一致のパーセンテージが決定されるという事が理解できる。したがって、例えば、下に論ずるように、野生型TRタンパク質配列より短い配列の配列一致（即ち、図2、図16Nを参照されたい）は、或る態様において、より短いその配列中のアミノ酸の数を用いて決定する。一致パーセントの算出において、相対重量は、挿入、欠失、置換などといった配列変異の諸表示には割り当てられない。

20

【0162】

或る態様では、一致のみが正の得点(+1)を得、ギャップを包含する全ての形式の配列変異は数値「0」が割り当てられるが、これは、配列類似性を算出するための、下記のような加重スケールまたはパラメータの必要性を取り除く。配列一致パーセントは例えば、完全に同一の残基の数を、アラインメントさせている領域内の「より短い」配列の総残基数で除し、100を掛けることによって算出できる。「より長い」配列は、アラインメントさせている領域中の実際の残基を殆ど持っている配列である。

30

【0163】

したがって、本発明に係る変異体TRタンパク質は、野生型TRタンパク質（即ち、図21）のアミノ酸配列より長くても短くてもよい。よって、好ましい態様では、本明細書に記載する配列の一部またはフラグメントが変異体TRタンパク質の定義に包含される。変異体TRタンパク質のフラグメントは、もしa) それらが少なくとも1個の抗原エピトープを共有し; b) 少なくとも指摘された相同性を持ち; c) そして好ましくは本明細書に定義した変異体TRの生物活性を持つならば、変異体TRタンパク質であると考えられる。

40

【0164】

好ましい態様では、後により詳細に記載するように、変異体TRタンパク質は、本明細書に概説したものに加えて、野生型TRと比較してさらなるアミノ酸変異を包含する。加えて、本明細書に概説するように、本明細書に記載の任意の変異は、任意の方法で合してさらなる新規な変異体TRタンパク質を形成することができる。

【0165】

さらに、図面に示すものより長い変異体TRタンパク質を、例えば本明細書に概説するようなエピトープまたは精製タグの付加、その他の融合配列の付加などにより製造できる。例えば、本発明に係る変異体TRタンパク質は、薬物動態学的目的のため、他の治療用タンパク質またはFcもしくは血清アルブミンといった他のタンパク質と融合させることができる

50

。例えば米国特許第5766883および5876969号を参照されたく、これらは共に引用によりここに含める。

【0166】

好ましい態様では、本発明に係る変異体TRタンパク質はヒトTRコンホマーである。本明細書中「コンホマー」とは、タンパク質バックボーンの3D構造は事実上同じであるがアミノ酸側鎖に著明な相違を持つタンパク質を意味する。即ち、本発明に係る変異体TRタンパク質とは、その集合のタンパク質の全てが1つのバックボーン構造を共有しており、それによって少なくとも1-3-5%相違する配列を持っている、コンホマー集合を定義するものである。したがって変異体TRタンパク質の三次元的バックボーン構造は実質的にヒトTRの三次元的バックボーン構造に対応する。この文脈中の「バックボーン」とは、非側鎖原子：窒素、カルボニル炭素および酸素、ならびに α 炭素、ならびに窒素および α 炭素に結合した水素を意味する。コンホマーであるとみなすためには、タンパク質は、ヒトTR構造から2オングストロームより大きく離れていない、好ましくは1.5オングストロームより大きく離れていない、そして特に好ましくは1オングストロームより大きく離れていないバックボーン原子を持っていないなければならない。一般に、これらの距離は二つの方法で測定できる。一つの態様では、可能性ある各々のコンホマーを結晶化し、その三次元構造を決定する。前述の方法はかなり冗長であるため、別法として、可能性ある各コンホマーの配列をPD Aプログラムでランし、これがコンホマーであるかどうかを決定する。

10

【0167】

これに代わる態様では、本発明に係る変異体TRタンパク質は、図21に列挙するTRタンパク質のいずれかのコンホマーであってよい。

20

【0168】

変異体TRタンパク質はまた、変異体TR核酸によりコードされているものとして同定できる。核酸の場合、核酸配列の全体的相同性はアミノ酸の相同性と同等であるが、遺伝的コードの縮重と、異なる生物のコドン偏倚を考慮する。したがって、核酸配列の相同性はタンパク質配列の相同性より低くても高くてもよいが、より低い相同性が好ましい。

【0169】

好ましい態様では、変異体TR核酸は変異体TRタンパク質をコードしている。当業者は理解できるであろうが、遺伝的コードの縮重のため、膨大な数の核酸が製造でき、それらの全てが本発明に係る変異体TRタンパク質をコードしている。したがって、特定のアミノ酸配列を同定したならば当業者は、変異体TRのアミノ酸配列を変化させない方法で単に1またはそれ以上のコドンの配列を修飾することによって、多数の異なる核酸を作製することができる。

30

【0170】

或る態様では、核酸の相同性をハイブリダイゼーション研究によって決定する。したがって、例えば、図21に示す核酸配列またはその相補物と高ストリンジェンシー下でハイブリダイズし、且つ変異体TRタンパク質をコードしている核酸が、変異体TR遺伝子であると考えられる。

【0171】

高ストリンジェンシー条件は当分野で既知であり、例えばManiatis et al., Molecular Cloning: A Laboratory Manual, 2d Edition, 1989、およびShort Protocols in Molecular Biology, ed. Ausubel, et al.を参照されたく、これらの両方を引用によりここに含める。ストリンジェント条件は配列依存的であり、異なる状況では相違する。長い配列は高温で特異的にハイブリダイズする。核酸のハイブリダイゼーションに対する広範な指針はTijssen, Techniques in Biochemistry and Molecular Biology--Hybridization with Nucleic Acid Probes, 「Overview of principles of hybridization and the strategy of nucleic acid assays」 (1993)に見出せる。一般にストリンジェント条件は、一定のイオン強度およびpHにおける特定配列の融点 (T_m) より約5-10°C低く選択する。 T_m は標的に対し相補的なプローブの50%が平衡時に標的配列とハイブリダイズする温度（一定のイオン強度、pHおよび核酸濃度において）である（標的配列は過剰に存在するため、 T_m におい

40

50

てはプローブの50%が平衡状態で占有される)。ストリンジェント条件は、pH7.0ないし8.3において塩濃度が約1.0Mナトリウムイオンより低い、典型的には約0.01ないし1.0Mのナトリウムイオン濃度(または他の塩)であり、短いプローブ(例えば10ないし50ヌクレオチド)のためには温度が少なくとも約30℃、長いプローブ(例えば50ヌクレオチドより大きい)のためには少なくとも約60℃であるような条件である。ストリンジェント条件はホルムアミドのような不安定化剤の添加によっても達成できる。

【0172】

別の態様では、よりストリンジェンシーの低いハイブリダイゼーション条件を使用し、例えば当分野で知られるように中等度または低ストリンジェンシー条件が使用できる。Maniatis and Ausubel(上記)およびTijssen(上記)を参照されたい。

10

【0173】

本発明に係る変異体TRタンパク質および核酸は組換えである。本明細書中使用する「核酸」とは、DNAもしくはRNAのいずれか、またはデオキシおよびリボヌクレオチドの両者を含む分子を指すことができる。この核酸はゲノムDNA、cDNAならびにセンスおよびアンチセンス核酸を包含するオリゴヌクレオチドを包含する。このような核酸は、生理的環境における係る分子の安定性と半減期を増大させるためにリボース-リン酸バックボーンに修飾を含むこともできる。

【0174】

この核酸は二本鎖、一本鎖であってよく、または二本鎖もしくは一本鎖配列の一部を含んでいてよい。当業者には理解できるであろうが、一本鎖(「Watson」という表現は他の鎖の配列(「Crick」)をも規定し、したがって図6に示す配列はこの配列の相補物をも包含する。本明細書中「組換え核酸」という語は、一般的にエンドヌクレアーゼによる核酸の操作により始めはインビトロで作製された、自然界では通常見出されない形態の核酸を意味する。したがって、直線型の単離された変異体TR核酸、または通常結合しないDNA分子をライゲーションすることによりインビトロで作製された発現ベクターは、いずれも本発明の目的のために組み換えられていると考えられる。いったん組換え核酸を製造し宿主細胞または生物内に再導入したならば、それは非組換え的に複製、即ち、インビトロ操作ではなく宿主細胞のインビボ細胞機構を利用して複製するが、いったん組換え的に作製されたこのような核酸は、その後非組換え的に複製を行っても、依然として本発明の目的のための組換えであると考えられる。

20

30

【0175】

同様に、「組換えタンパク質」とは、組換え技術を用いて、即ち上述の組換え核酸の発現によって製造したタンパク質である。組換えタンパク質は少なくとも1またはそれ以上の性質により天然に存在するタンパク質と区別できる。例えば、このタンパク質はその野生型宿主に通常付随するタンパク質および化合物の幾らかまたは全てを含まないよう単離または精製でき、よって実質上純粋であり得る。例えば、単離されたタンパク質は、好ましくは与えられた試料中の総タンパク質の少なくとも約0.5重量%、より好ましくは少なくとも約5重量%を構成する、天然状態で通常付随する物質の少なくとも幾らかを含まない。実質上純粋なタンパク質は総タンパク質の少なくとも約75重量%を含み、少なくとも約80%が好ましく、少なくとも約90%が特に好ましい。この定義は、或る生物に由来する変異体TRタンパク質を、異なる生物または宿主細胞において製造することを包含する。これとは別に当該タンパク質は、このタンパク質が増大した濃度レベルで製造されるよう誘導プロモーターまたは高発現プロモーターを使用することにより、通常観察される濃度よりはるかに高い濃度で製造することができる。さらに、本明細書に概説する全ての変異体TRタンパク質は、下に論ずるように、アミノ酸置換、挿入および欠失を含むため(置換が好ましい)、天然に通常見出されない形態である。

40

【0176】

本明細書に概説し図面に示す変異体TR配列のアミノ酸配列変異体もまた本発明に係る変異体TRタンパク質の定義に包含される。即ち、この変異体TRタンパク質はヒトTRに比してさらなる可変位置を含んでいてよい。これらの変異体は3つのクラス:置換、挿入または欠

50

失変異体のうち1またはそれ以上に該当する。これらの変異体は普通、カセットまたはPCR突然変異誘発またはその他の当分野で周知の技術を用いる、該変異体をコードしているDNAを生成するための、変異体TRタンパク質をコードしているDNAにおけるヌクレオチドの位置特異的突然変異誘発によって製造し、その後上に概説するように組換え細胞培養において該DNAを発現させる。しかしながら約100-150までの残基を持つ変異体TRタンパク質フラグメントは、確立した技術を用いるインビトロ合成によって製造できる。アミノ酸配列変異体は、変異体TRタンパク質アミノ酸配列の天然に存在するアリルまたは種間変異とそれらを隔てる特徴である、前もって決定されている変異の性質によって特徴付けられている。この変異は典型的には天然に存在する類似体と同じ質的生物活性を示すが、下記により詳細に述べるように、改変された性質を持つ変異体を選択することもできる。

10

【0177】

アミノ酸配列変異を導入するための部位または領域は前もって決定しておくが、突然変異自体を前もって決定しておく必要はない。例えば、所定の部位での突然変異の出来を最適なものとするために、標的コドンまたは領域でランダム突然変異誘発を実施し、発現された変異体TRタンパク質を所望の活性の最適な組み合わせについてスクリーニングすることができる。既知配列を持つDNAにおいて前もって定められた部位に置換突然変異を作製する技術は周知であり、例えばM13プライマー突然変異誘発およびPCR突然変異誘発がある。突然変異体のスクリーニングは変異体TRタンパク質活性の検定を用いて行う。

【0178】

アミノ酸置換は典型的には単一残基の置換であり、挿入は通常約1ないし20アミノ酸であるが、かなり多くの挿入も許容できる。欠失は約1ないし約20残基の範囲であるが、幾つかの場合には欠失はずっと大きくなり得る。

20

【0179】

置換、欠失、挿入またはこれらの任意の組み合わせを用いて最終誘導体に到達できる。一般にこれらの変化は当該分子の変化を最小とするため少数アミノ酸について行う。しかしながら或る状況では、より大きな変化が許容できる。変異体TRタンパク質の性質に小さな変化を所望する場合、置換は一般に以下の表に従って行う。

【0180】

表1

【表1】

30

もとの残基	例示的置換
Ala	Ser
Arg	Lys
Asn	Gln, His
Asp	Glu
Cys	Ser, Ala
Gln	Asn
Glu	Asp
Gly	Pro
His	Asn, Gln
Ile	Leu, Val
Leu	Ile, Val
Lys	Arg, Gln, Glu
Met	Leu, Ile
Phe	Met, Leu, Tyr
Ser	Thr
Thr	Ser
Trp	Tyr
Tyr	Trp, Phe
Val	Ile, Leu

40

【0181】

表1に示したものより同類性の低い置換を選択することにより、機能または免疫学的同一

50

性に実質的な変化を創出する。例えば、変化領域におけるポリペプチドバックボーンの構造、例えば α ヘリックスまたは β シート構造；標的部位の分子の電荷または疎水性；または側鎖の嵩、に、より著しい影響を及ぼす置換を創出できる。当該ポリペプチドの性質に最大の変化を導くことが一般的に予想される置換は、(a) 親水性残基、例えばセリンまたはスレオニンを疎水性残基、例えばロイシン、イソロイシン、フェニルアラニン、バリンまたはアラニンに置換する（またはその逆）；(b) システインまたはプロリンを他の何らかの残基に置換する（またはその逆）；(c) 正電荷を持つ側鎖を有する残基、例えばリジン、アルギニン、またはヒスチジンを、負電荷を持つ残基、例えばグルタミン酸またはアスパラギン酸に置換する（またはその逆）；または(d) 嵩高い側鎖を持つ残基、例えばフェニルアラニンを、側鎖を持たない残基、例えばグリシンに置換する（またはその逆）、置換である。 10

【0182】

変異体は典型的には元の変異体TRタンパク質と同じ質的生物活性を示し、同じ免疫反応を誘起するが、必要に応じて変異体TRタンパク質の性格を修飾するように変異体を選択することもある。これとは別に該変異体は、変異体TRタンパク質の生物活性が変化するように設計できる。例えば、グリコシル化部位を変化させまたは除去することができる。同様に、生物学的機能を変化させることができ、例えば幾つかの例では、より強力なまたはより強力でないTR活性を持つことが望ましいかも知れない。

【0183】

本発明に係る変異体TRタンパク質および核酸は幾つかの方法で製造できる。個々の核酸およびタンパク質は、当分野で知られているように、そして下に概説するように製造できる。これとは別に、変異体TRタンパク質のライブラリーを試験のために作製できる。 20

【0184】

好ましい態様では、変異体TRタンパク質の集合またはライブラリーを確率分布表から作製する。本明細書に概説するように、PDA、配列アラインメント、SCMF計算のような力場計算等を包含する、確率分布表を作製する様々な方法がある。加えて、確率分布は、ライブラリー中に観察される突然変異頻度の尺度としての、各位置についての情報エントロピー評点の作製に使用できる。

【0185】

この態様では、このリスト中の各々の可変位置にある各アミノ酸残基の頻度を同定する。頻度には限界が設定でき、ここではカットオフより低い変異体頻度を0に設定する。このカットオフは好ましくは1%、2%、5%、10%または20%であり、10%が特に好ましい。次いでこれらの頻度を変異体TRライブラリー中に組み入れる。即ち、上記のように、これらの可変位置を集め、可能な全ての組み合わせを作製するが、但しライブラリーを「満たす」アミノ酸残基は頻度をベースとして利用する。したがって、頻度に基づかないライブラリーでは、5個の可能な残基を持つ可変位置は、第一の可能な残基を持つ可変位置を含むタンパク質20%、第二を持つもの20%、などを有する。しかしながら頻度に基づくライブラリーでは、それぞれ10%、15%、25%、30%および20%の頻度を持つ5個の可能な残基を有する可変位置は、第一の可能な残基を持つ可変位置を含むタンパク質10%、第二の残基を持つタンパク質15%、第三を持つタンパク質25%等を有するであろう。当業者には理解できるであろうが、実際の頻度はそのタンパク質を実際に作製するために利用する方法に依存し、例えば、正確な頻度はそのタンパク質を合成する際に取得可能となり得る。しかしながら、以下に概説する頻度に基づくプライマー系を使用する場合、各位置における実際の頻度は、以下に概説するように、変化するであろう。 30 40

【0186】

当業者には理解できるであろうが、そして本明細書に概説するように、確率分布表は様々な方法で作製できる。本明細書に概説する方法に加えて、自己無撞着平均場（SCMF）法を、確率表の直接作製に使用できる。SCMFは回転異性体相互作用の平均場の描写を用いてエネルギーを計算する、コンピューターによる決定論的方法である。このようにして作製した確率表を用いて本明細書に記載のライブラリーが作製できる。SCMFは3つのやり方で利 50

用できる。アミノ酸の頻度および各アミノ酸についての回転異性体を各位置において列挙する；確率をSCMFから直接決定する（Delarue et al. Pac. Symp. Biocomput. 109-21 (1997) (引用によりここに含める) を参照されたい）。加えて、非常に可変的な位置および非可変位置が同定できる。これに代わり、別の方法を利用して、配列空間の検索中にどの配列がとばされたかを決定し；SCMFを利用してその配列についての正確なエネルギーを得；次いでこのエネルギーを、等級付けに利用し、配列の等級順リストを作る（Monte Carlo配列リストに類似）。しかる後、各位置のアミノ酸の頻度を示す確率表がこのリストから算出できる（Koehl et al., J. Mol. Biol. 239:249 (1994); Koehl et al., Nat. Struct. Biol. 2:163 (1995); Koehl et al., Curr. Opin. Struct. Biol. 6:222 (1996); Koehl et al., J. Mol. Biol. 293:1183 (1999); Koehl et al., J. Mol. Biol. 293:1161 (1999); Lee J. Mol. Biol. 236:918 (1994); およびVasquez Biopolymers 36:53-70 (1995) ; これらは全て引用によりここに含める）。同様の方法には、OPLS-AA (Jorgensen, et al., J. Am. Chem. Soc. (1996), v 118, pp 11225-11236; Jorgensen, W.L.; BOSS, Version 4.1; Yale University: New Haven, CT (1999)); OPLS (Jorgensen, et al., J. Am. Chem. Soc. (1988), v 110, pp 1657ff; Jorgensen, et al., J. Am. Chem. Soc. (1990), v 112, pp 4768ff); UNRES (United Residue Forcefield; Liwo, et al., Protein Science (1993), v 2, pp1697-1714; Liwo, et al., Protein Science (1993), v 2, pp1715-1731; Liwo, et al., J. Comp. Chem. (1997), v 18, pp849-873; Liwo, et al., J. Comp. Chem. (1997), v 18, pp874-884; Liwo, et al., J. Comp. Chem. (1998), v 19, pp259-276; Forcefield for Protein Structure Prediction (Liwo, et al., Proc. Natl. Acad. Sci. USA (1999), v 96, pp5482-5485); ECEPP/3 (Liwo et al., J Protein Chem 1994 May;13(4):375-80); AMBER 1.1 force field (Weiner, et al., J. Am. Chem. Soc. v106, pp765-784); AMBER 3.0 force field (U.C. Singh et al., Proc. Natl. Acad. Sci. USA. 82:755-759); CHARMM and CHARMM22 (Brooks, et al., J. Comp. Chem. v4, pp 187-217); cvff3.0 (Dauber-Osguthorpe, et al., (1988) Proteins: Structure, Function and Genetics, v4, pp31-47); cff91 (Maple, et al., J. Comp. Chem. v15, 162-182) が包含され（但しこれらに限定される訳ではない）；さらに、the DISCOVER (cvff and cff91) およびAMBER力場がthe INSIGHT molecular modeling package (Biosym/MSI, San Diego California) で使用され、そしてHARMMがthe QUANTA molecular modeling package (Biosym/MSI, San Diego California) で使用されている。

【0187】

さらに、本明細書に概説するように、確率分布表を作製する好ましい方法は、配列アラインメントプログラムの使用による。加えて確率表は配列アラインメントとコンピューターアプローチの組み合わせによって得られる。例えば、相同的配列のアラインメントに見出されるアミノ酸をコンピューター計算の結果に加えることができる。好ましくは野生型アミノ酸を、これがコンピューター計算に見出されない場合に確率表に加えることができる。

【0188】

理解できるであろうが、可変位置および/または可変位置にある残基を再結合して創出した変異体TRライブラリーは等級順リストに無いかも知れない。幾つかの態様では、このリスト全体を作製し試験できる。これに代わり好ましい態様では、この変異体TRライブラリーもまた等級順リストの形態である。これは、ライブラリーのサイズが実験的に作製するには依然として大きすぎるという事を包含する幾つかの理由のため、または予想目的のために行われる。これは幾つかの方法で行うことができる。

【0189】

或る態様では、ライブラリーは、ライブラリーの成員を等級付けるためのPDAの評点機能を用いて等級付ける。別法として統計学的方法が使用できる。例えば、ライブラリーは頻度評点によって等級付けることができる。即ち、殆どの高頻度残基を含むタンパク質を高い等級とするなどである。これは、評点数値を求めるため各可変位置での頻度を加えるまたは掛けることにより行うことができる。同様に、ライブラリーの異なる位置を加重し次

いでそのタンパク質の評点を付けることができるが、例えば或る一定の残基を含むものを恣意的に等級付けることができる。

【0190】

好ましい態様では、変異体TRライブラリーの異なるタンパク質成員を化学合成できる。この事は、設計したタンパク質が短い、好ましくは150アミノ酸長より短い、さらに好ましくは100アミノ酸長より短い、そして特に好ましくは50アミノ酸長より短い場合にとりわけ有用であるが、当分野で知られるように、長いタンパク質を化学的または酵素的に製造することもできる。例えば、Wilken et al, Curr. Opin. Biotechnol. 9:412-26 (1998)を参照されたく、これは引用によりここに含める。

【0191】

好ましい態様では、とりわけ長いタンパク質または大試料が所望であるタンパク質のための好ましい態様では、成員の配列をコードしているDNAのような核酸を創出するためにこのライブラリー配列を使用し、次いでそれを宿主細胞中にクローニングし、発現させ、所望により検定する。したがって、各成員タンパク質配列をコードしている核酸、とりわけDNAを作製することができる。これは既知の方法を利用して行う。コドン、適当な発現ベクターおよび適当な宿主細胞の選択は幾つかの因子に依存して変化し、必要に応じて容易に最適化できる。

10

【0192】

好ましい態様では、U.S.S.N.09/927790 (引用によりここに含める) に一般的に記載されているように、プールしたオリゴヌクレオチドによるマルチプルPCR反応を行う。この態様では、完全長遺伝子に対応する部分重複オリゴヌクレオチドを合成する。さらに、これらのオリゴヌクレオチドは各変異体の位置または部分集合にある異なるアミノ酸の全てを表し得る。

20

【0193】

好ましい態様では、これらのオリゴヌクレオチドを等しい比率でプールし、マルチプルPCR反応を行って、該ライブラリーにより定義される突然変異の組み合わせを含む完全長配列を作り出す。加えて、これは誤りがちなPCR法を用いて実施できる。

【0194】

好ましい態様では、異なるオリゴヌクレオチドを確率分布表に対応する相対量で加える。こうしてマルチプルPCR反応は、所望の突然変異の組み合わせを所望の比率で有する完全長配列を生成する。

30

【0195】

必要なオリゴヌクレオチドの総数は、突然変異を受ける位置の数およびこれらの位置にあると考えられる突然変異の数の関数である：

$$(\text{定位置に対するオリゴの数}) + M1 + M2 + M3 + \dots + Mn = (\text{必要なオリゴの総数})$$

[式中、Mnはこの配列の位置nにあると考えられる突然変異の数である]。

【0196】

好ましい態様では、各々の部分重複するオリゴヌクレオチドは、変化させるべき位置をただ1個含んでおり；別の態様では、そうなるには変異体位置が近すぎるので、オリゴヌクレオチド1個当たり複数の変異体を使用してあらゆる可能性を有する完全な組換えをさせる。即ち、各オリゴは突然変異を受ける1個の位置のためのコドン、または突然変異を受ける1以上の位置のためのコドンを含むことができる。突然変異を受ける複数の位置は、オリゴの長さが非実用的となるのを防ぐため、配列中で近接していなければならない。或るオリゴヌクレオチド上の複数の突然変異位置を目的として、突然変異の特定の組み合わせを、その組み合わせをコードしているオリゴヌクレオチドを包含または除外することによってそのライブラリーに包含またはそこから排除することができる。例えば、本明細書に論ずるように、可変領域の間には相関があり得る。即ち、位置Xが或る残基である場合、位置Yは特定の残基でなければならない（またはそうであってはならない）。これらの可変位置の集合は、本明細書では時に「クラスター」と呼ぶ。クラスターが近接した残基に含まれ、よって1個のオリゴヌクレオチドプライマー上に存在する場合、このクラスター

40

50

ーは「良好な」相関関係へと向かい、ライブラリーの有効性を低減させ得る劣悪な相関を排除することができる。しかしながら、クラスターの残基が配列内で遠く離れており、よって合成のための異なるオリゴヌクレオチド上に存在する場合、該残基を「良好な」相関関係に置くか、またはそれらを可変残基として全て排除するのが望ましいであろう。これとは別の態様では、クラスター突然変異が常に一諸に現れるよう、ライブラリーを幾つかの工程で作製できる。この方法、即ち突然変異のクラスターを同定し、それらを同じオリゴヌクレオチド上に位置させるか、またはライブラリーからそれらを排除する方法、またはクラスターを温存する幾つかの工程によるライブラリーの作製は、適正に折り畳まれたタンパク質を有する実験的ライブラリーをかなり高品質なものとし得る。クラスターの同定は幾つかの方法で、例えば既知のパターン認識法、または突然変異の発生頻度の比較を用いて、または実験的に製造しようとする配列のエネルギー解析（例えば、相互作用のエネルギーが高い場合、その位置は相関している）によって、実施できる。これらの相関は、位置相関（例えば、可変位置1と2が常に一諸に変化しまたは決して一諸に変化しない）または配列相関（例えば、1位に残基Aがある場合、2位には常に残基Bがある）であってよい。Pattern discovery in Biomolecular Data: Tools, Techniques, and Applications; edited by Jason T.L. Wang, Bruce A. Shapiro, Dennis Shasha. New York: Oxford University, 1999; Andrews, Harry C. Introduction to mathematical techniques in pattern recognition; New York, Wiley-Interscience [1972]; Applications of Pattern Recognition; Editor, K.S. Fu. Boca Raton, Fla. CRC Press, 1982; Genetic Algorithms for Pattern Recognition; edited by Sankar K. Pal, Paul P. Wang. Boca Raton: CRC Press, c1996; Pandya, Abhijit S., Pattern recognition with neural networks in C++ / Abhijit S. Pandya, Robert B. Macy. Boca Raton, Fla.: CRC Press, 1996; Handbook of pattern recognition & computer vision / edited by C.H. Chen, L.F. Pau, P.S.P. Wang. 2nd ed. Singapore; River Edge, N.J.: World Scientific, c1999; Friedman, Introduction to Pattern Recognition: Statistical, Structural, Neural, and Fuzzy Logic Approaches; River Edge, N.J.: World Scientific, c1999, Series title: Series in machine perception and artificial intelligence; vol. 32（これらは全て引用によりここに含める）を参照されたい。さらに、コンセンサスモチーフの探索に用いられるプログラムもまた使用できる。

10

20

30

【0197】

加えて、オリゴヌクレオチドの設計を変えることにより、即ち、そのオリゴヌクレオチド（プライマー）がどこで開始および停止するか（例えばどこでその配列が「切断」されるか）を決定することにより、相関およびシャフリングを固定または最適化できる。オリゴの開始および停止部位は、単一オリゴヌクレオチド中に現れるクラスターの数を最大とするよう設定でき、それにより、高評点の配列でそのライブラリーを質的に向上させることができる。異なるオリゴヌクレオチド開始および停止部位の選択肢は、単一のオリゴ上にあるクラスター数に従って、または予想した配列ライブラリーに一致する得られた配列のパーセンテージに従って、コンピューターにより作製および等級付けできる。

【0198】

必要なオリゴヌクレオチドの総数は、複数の突然変異可能な位置が単一オリゴヌクレオチドによってコードされている場合に増加する。アニーリングされた領域は、一定不変の領域、即ちレファレンス配列を有する領域である。

40

【0199】

コドンの挿入または欠失を持つオリゴヌクレオチドを使用して異なる長さのタンパク質を発現するライブラリーを創出できる。とりわけ、挿入または欠失を求めるコンピューターによる配列スクリーニングは、異なる長さのタンパク質を規定する二次ライブラリーを産み、このタンパク質はプールされた異なる長さのオリゴヌクレオチドのライブラリーにより発現され得る。

【0200】

好ましい態様では、変異体TRライブラリーは、ファミリー（例えば変異体の集合）のシャ

50

フリリングによって行う。即ち、誤りがちなPCRによってまたはこれによらずに上位配列の集合（もし等級順リストを使用するならば）をシャフルすることができる。この文脈中の「シャフリリング」とは、一般にランダムなやり方での関連配列の再結合を意味する。これは、米国特許Nos. 5,830,721; 5,811,238; 5,605,793; 5,837,458およびPCT US/19256（これらは全て引用によりその全内容をここに含める）に定義および例示されている「シャフリリング」を包含できる。この配列の集合は人工的集合、例えば確率表（例えばSCMFを使用して作製されたもの）またはMonte Carlo集合由来のものとする 것도できる。同様に、「ファミリー」とは、上位の10および下位の10配列、上位の100配列などとする 것도できる。これもまた誤りがちなPCRを用いて実施できる。

【0201】

10

したがって好ましい態様では、本明細書に記載のコンピューター的方法を用いるin silicoシャフリリングを行う。即ち、2個のライブラリーまたは2個の配列から始まって、配列のランダムな再結合を作製し評価することができる。

【0202】

好ましい態様では、変異体TRライブラリーを作製するために誤りがちなPCRを行う。米国特許Nos. 5,605,793, 5,811,238および5,830,721（これらは全て引用によりここに含める）を参照されたい。これは最適配列について、またはライブラリーもしくは他の何らかの人工的集合もしくはファミリーの上位成員について実施できる。この態様では、一次ライブラリーのコンピュータースクリーニングで見出された最適配列の遺伝子が合成できる。次いで、該ライブラリーの変異体位置に突然変異をコードしているオリゴヌクレオチド（バイアスオリゴヌクレオチド）の存在下で、誤りがちなPCRを最適配列遺伝子上で実施する。このオリゴヌクレオチドの添加は、ライブラリーに突然変異を組み込む傾向のあるバイアスを創出する。別法として、或る突然変異をコードしているオリゴヌクレオチドのみを用いてライブラリーにバイアスをかけることもできる。

20

【0203】

好ましい態様では、誤りがちなPCRによる遺伝子シャフリリングを、バイアスオリゴヌクレオチドの存在下に最適配列の遺伝子上で実施し、変異体TRライブラリー中に見出される突然変異の比率を反映するDNA配列ライブラリーを創出することができる。バイアスオリゴヌクレオチドの選択は様々な方法で実施できる。これらはその頻度に基づいて選択できる。即ち、高突然変異頻度の位置をコードしているオリゴヌクレオチドが使用でき；そうではなく、多様性が増大するよう最も変化し易い位置を含むオリゴヌクレオチドが使用でき；二次ライブラリーを等級付ける場合には幾つかの上位評点位置をバイアスオリゴヌクレオチドの作製に使用でき；ランダムな位置を選択でき；数個の上位評点および数個の低評点オリゴヌクレオチドが選択できる、などである。重要な事は、好ましい可変位置および配列に基づいて新しい配列を作製することである。

30

【0204】

好ましい態様では、U.S.S.N. 09/927,790（引用によりここに含める）に一般的に記載されるような野生型遺伝子またはその他の遺伝子を用いるPCRが利用できる。この態様では、開始遺伝子を使用し、一般にこの遺伝子は通常野生型遺伝子である（但しそれが必要であるわけではない）。幾つかの例では、これは全体として最適化された配列またはリスト中の他の何らかの配列または、例えば異なる生物由来の相同的配列をアラインメントすることで得られたコンセンサス配列をコードしている遺伝子とすることができる。この態様では、変異体位置に対応し且つライブラリー中の異なるアミノ酸を含むオリゴヌクレオチドを使用する。当分野で知られるように、末端にPCRプライマーを用いてPCRを行う。これには2つの利点がある。第一は、一般にこれは少ないオリゴヌクレオチドしか必要とせず、誤りがより少なくなる。加えて、野生型遺伝子を使用する場合、それを合成する必要が無いという実験上の有利性がある。

40

【0205】

さらに、図面に例示したような、利用できる他の幾つかの技術がある。好ましい態様では、PCR生成物のライゲーションを実施する。

50

【0206】

好ましい態様では、変異体TRライブラリーに対して多様なさらなる工程を実施することができる。例えばさらなるコンピューター処理を行い、異なる変異体TRライブラリーが再結合でき、または異なるライブラリー由来のカットオフを組み合わせることができる。好ましい態様では、変異体TRライブラリーを再度コンピューター操作して、さらなる変異体TRライブラリーを作ることができる（時にこれを本明細書では「第三のライブラリー」と称する）。例えば、任意の変異体TRライブラリー配列を、第一ライブラリー中の変化した位置の幾つかまたは全てを凍結または固定することにより、二巡目のPDAのために選択できる。別法として、最後の確率分布表に見られる変化のみを許容する。或いは、確率表のストリンジェンシーを、包含についてのカットオフを増大または低下させることにより、変化させることができる。同様に、変異体TRライブラリーは、一巡目の後に実験的に再結合させる事ができる。例えば、第一スクリーニング由来の最良遺伝子/遺伝子群を選択し、再度遺伝子アセンブリーを行うことができる（下に概説する技術、マルチプルPCR、誤りがちなPCR、シャフリングなどを利用）。

10

【0207】

これに代わり、1またはそれ以上の良好遺伝子（群）由来のフラグメントが幾つかの位置で確率を変化させる。これは、コンピューターおよび実験的スクリーニングの一巡目に見出される配列空間の領域に検索を偏倚させる。

【0208】

好ましい態様では、組み合わせた異なる変異体TRライブラリーから第三ライブラリーが作製できる。例えば、本明細書に概説するように、第一の変異体TRライブラリー由来の確率分布表を、コンピューターまたは実験で作製し再結合する。PDA変異体TRライブラリーを配列アラインメント変異体TRライブラリーと合し、再結合（やはりコンピューターまたは実験によって）させるか、または各々のカットオフのみを合わせて新しい第三ライブラリーを作製する。幾つかのライブラリーからの上位配列を再結合できる。ライブラリー上位からの配列をライブラリー下位からの配列と合してより広範な試料配列空間を作製し、または、ライブラリー上位から遠い配列のみを合する。タンパク質の異なる部分を解析した変異体TRライブラリーを、そのタンパク質の合した部分を処理する第三ライブラリーと合することができる。

20

【0209】

好ましい態様では、第三ライブラリーを変異体TRライブラリー中の相関を用いて作製できる。即ち、第一の可変位置にある残基は第二可変位置の残基と相関している（またはさらなる位置にある残基と同様に相関している）かも知れない。例えば、第一の残基がXであれば第二の残基はYでなければならないというように、2個の可変位置が立体的または静電的に相互作用しているかも知れない。これは正の相関であっても負の相関であってもよい。

30

【0210】

候補変異体タンパク質または候補変異体ライブラリー成員をコードしている本発明に係る核酸を使用して、様々な発現ベクターを作製する。この発現ベクターは、自己複製する染色体外ベクターであっても、また、宿主ゲノム中に統合するベクターであってもよい。一般にこれらの発現ベクターは、ライブラリーのタンパク質をコードしている核酸と機能的に連結した転写および翻訳調節核酸を含む。「調節配列」という語は、特定の宿主生物において機能的に連結したコード配列の発現に必要とされるDNA配列を指す。例えば原核生物のために好適な調節配列は、プロモーター、所望によりオペレーター配列、およびリボソーム結合部位を包含する。真核生物細胞は、プロモーター、ポリアデニル化シグナル、およびエンハンサーを利用することが知られている。

40

【0211】

核酸は、それが別の核酸配列と機能的関係に配置される場合、「機能的に連結」している。例えば、プレ配列のためのDNAまたは分泌リーダーは、或るポリペプチドの分泌に参加するプレタンパク質としてこれが発現されるならば、そのポリペプチドのDNAと機能的に

50

連結しており；プロモーターまたはエンハンサーは、これが配列の転写に影響を及ぼすならば、コード配列と機能的に連結しており；または、リボソーム結合部位は、翻訳を促進するように位置しているならば、コード配列と機能的に連結している。一般に、「機能的に連結」とは、連結しているDNA配列が隣接しており、そして分泌リーダーの場合は隣接し且つ読み取り相にある。しかしながらエンハンサーは隣接していなくともよい。連結は都合の良い制限部位でのライゲーションにより達成する。そのような部位が無い場合は、常法に従い合成オリゴヌクレオチドアダプターまたはリンカーを使用する。当業者には理解できるであろうが、転写および翻訳調節核酸は一般に、そのライブラリータンパク質を発現するのに用いられる宿主細胞に適している。例えば、好ましくはBacillus由来の転写および翻訳調節核酸配列を使用してBacillus中でライブラリータンパク質を発現させる。様々な宿主のための、数多くの種類の適当な発現ベクター、および適当な調節配列が当分野で知られている。

10

【0212】

一般に転写および翻訳調節配列は、プロモーター配列、リボソーム結合部位、転写開始および停止配列、翻訳開始および停止配列、ならびにエンハンサーまたはアクチベーター配列を包含するが、これらに限定される訳ではない。好ましい態様では、調節配列はプロモーターならびに転写開始および停止配列を包含する。

【0213】

プロモーター配列は構成的および誘導的プロモーター配列を包含する。プロモーターは天然に存在するプロモーター、ハイブリッドまたは合成プロモーターを包含できる。1以上のプロモーターの要素を合しているハイブリッドプロモーターもまた当分野で知られており、本発明において有用である。

20

さらに、発現ベクターはさらなる要素を含み得る。例えば、発現ベクターは2つの複製系を持つことができ、したがってこれを2つの生物、例えば発現のための哺乳動物または昆虫細胞において、そしてクローニングおよび増幅のための原核生物宿主において維持することを可能にする。さらに、発現ベクターを統合するため、この発現ベクターは、宿主細胞ゲノムに相同的な少なくとも1個の配列、および好ましくは発現構築物に隣接する2個の相同的配列を含む。この統合ベクターは、ベクターに含まれるための適当な相同配列を選択することにより、宿主細胞中の特定の座を目指すことができる。ベクター統合のための構築物ならびに適当な選択およびスクリーニングプロトコルは当分野で周知であり、例えばMansour et al., Cell, 51:503 (1988) およびMurray, Gene Transfer and Expression Protocols, Methods in Molecular Biology, Vol. 7 (Clifton: Humana Press, 1991)に記載されている。

30

【0214】

さらに、好ましい態様では、発現ベクターは、発現ベクターを含む形質転換された宿主細胞の選択を可能にする選択遺伝子を含み、特に哺乳動物細胞の場合には、該ベクターを含まない細胞は一般に死滅することからベクターの安定性を確実とする。選択遺伝子は当分野で周知であり、使用する宿主細胞と共に変わる。本明細書中「選択遺伝子」とは、選択剤に対する耐性を付与する遺伝子産物をコードしている任意の遺伝子を意味する。好適な選択剤は、ネオマイシン（またはその類似体G418）、プラスティシディンS、ヒスチニドールD、ブレオマイシン、ピューロマイシン、ハイグロマイシンB、およびその他の薬物を包含するが、これらに限定される訳ではない。

40

【0215】

好ましい態様では、遺伝子発現のレベルを増大させるため、発現ベクターはRNAスプライシング配列を、発現させようとする遺伝子上流または下流に含む。Barret et al., Nucleic Acids Res. 1991; Groos et al., Mol. Cell. Biol. 1987; およびBudiman et al., Mol. Cell. Biol. 1988を参照されたい。

【0216】

好ましい発現ベクター系は一般的にMann et al., Cell, 33:153-9 (1993); Pear et al., Proc. Natl. Acad. Sci. U.S.A., 90(18):8392-6 (1993); Kitamura et al., Proc. Na

50

tl. Acad. Sci. U.S.A., 92:9146-50 (1995); Kinsella et al., Human Gene Therapy, 7:1405-13; Hofmann et al., Proc. Natl. Acad. Sci. U.S.A., 93:5185-90; Choate et al., Human Gene Therapy, 7:2247 (1996); PCT/US97/01019およびPCT/US97/01048、ならびにこれらに引用されている文献（これらは全て引用によりここに含める）に記載されているような、レトロウイルスベクター系である。

【0217】

本発明に係る候補変異体ライブラリータンパク質は、核酸、好ましくはライブラリータンパク質をコードしている核酸を含む発現ベクターで形質転換された宿主細胞を、そのライブラリータンパク質の発現を誘導または惹起する適当な条件下で培養することによって生産する。候補変異体ライブラリータンパク質発現にとって適当な条件は発現ベクターおよび宿主細胞の選択によって変わり、日常的な実験を介して当業者が容易に確認できる。例えば、発現ベクター中での構成的プロモーターの使用は宿主細胞の成長と増殖の最適化を必要とし、一方誘導的プロモーターの使用は誘導のために適当な成長条件を必要とする。加えて、幾つかの態様では、収穫のタイミングが重要である。例えば、昆虫細胞発現に使用するバキュロウイルス系は溶菌ウイルスであり、したがって収穫時期の選択が生成物の収量にとって決定的となり得る。

10

【0218】

当業者には理解できるであろうが、本発明で使用する細胞の種類は広範に変わり得る。基本的には、酵母、細菌、始原細菌、真菌、および昆虫、植物、ならびに哺乳動物細胞を包含する動物細胞を包含する、多岐にわたる適当な宿主細胞が使用できる。Drosophila melanogaster細胞、Saccharomyces cerevisiaeおよびその他の酵母、E.coli、Bacillus subtilis、SF9細胞、C129細胞、293細胞、Neurospora、BHK、CHO、COS、およびHeLa細胞、線維芽細胞、Schwanomaセルライン、不死化した哺乳動物骨髄系およびリンパ系セルライン、Jurkat細胞、肥満細胞およびその他の内分泌および外分泌細胞、ならびに神経細胞は特に興味深い。ATCCセルラインカタログ（これは引用によりここに含める）を参照されたい。加えて、例えば当分野で周知のファージディスプレイ系における二次ライブラリーの発現が、とりわけこの二次ライブラリーがランダムなペプチドを含む場合には好ましい。或る態様では、細胞を遺伝子操作でき、これは即ち例えば標的分子を含むよう、外因性核酸を含むという事である。

20

【0219】

好ましい態様では、候補変異体タンパク質または候補変異体ライブラリータンパク質を哺乳動物細胞中で発現させる。任意の哺乳動物細胞が使用でき、マウス、ラット、霊長類およびヒト細胞が特に好ましいが、当業者に理解できるように、シェードタイピングによるこの系の修飾によって、全ての真核生物細胞、好ましくは高等真核生物細胞が使用できるようになる。下により詳細に述べることであるが、細胞がランダムライブラリー成員の存在下で選択可能な表現型を示すよう、スクリーニングを設定する。下により詳細に述べることであるが、好適なスクリーニングを設計して、その細胞内部にライブラリー成員が存在する帰結として変化した表現型を示す細胞を選択できる限り、多彩な疾病状態に関連する細胞型が特に有用である。

30

【0220】

したがって、好適な哺乳動物細胞型は、全ての種類の腫瘍細胞（とりわけ黒色腫、骨髄性白血病、肺癌、乳癌、卵巣癌、結腸癌、腎臓癌、前立腺癌、膵臓癌および精巣癌）、心筋細胞、内皮細胞、上皮細胞、リンパ球（T細胞およびB細胞）、肥満細胞、好酸球、血管内皮細胞、肝細胞、単核白血球を包含する白血球、幹細胞、例えば造血、神経、皮膚、肺、腎臓、肝臓および筋細胞幹細胞（分化および非分化因子についてのスクリーニングに使用）、破骨細胞、軟骨細胞およびその他の結合組織細胞、ケラチノサイト、メラノサイト、肝臓細胞、腎臓細胞、および脂肪細胞を包含するが、これらに限定される訳ではない。好適な細胞はまた、Jurkat T細胞、NIH3T3細胞、CHO、Cos、等を包含する既知の研究用細胞を包含するが、これらに限定されない。引用によりここに含めるATCCセルラインカタログを参照されたい。

40

50

【0221】

哺乳動物発現系もまた当分野で知られており、レトロウイルス系を包含する。哺乳動物プロモーターは、哺乳動物RNAポリメラーゼと結合でき、ライブラリータンパク質のコード配列からmRNAへの下流(3')方向への転写を開始できる任意のDNA配列である。プロモーターは、通常コード配列の5'末端の近位に位置する転写開始領域、および、この転写開始部位の上流に位置する25-30塩基対を用いるTATAボックスを有する。TATAボックスはRNAポリメラーゼIIに指令して正しい部位でRNA合成を開始させると考えられる。哺乳動物プロモーターはさらに、典型的にはTATAボックスの上流100ないし200塩基対以内に位置する上流プロモーター要素(エンハンサー要素)を含む。上流プロモーター要素は転写を開始する速度を決定し、いずれの向きにも作用できる。哺乳動物ウイルス遺伝子由来のプロモーターが哺乳動物プロモーターとして特に有用であり、何故ならこのウイルス遺伝子はしばしば高度に発現され、広範な宿主範囲を有するためである。例には、SV40初期プロモーター、マウス乳癌ウイルスLTRプロモーター、アデノウイルス主要後期プロモーター、単純ヘルペスウイルスプロモーター、およびCMVプロモーターがある。

10

【0222】

典型的には、哺乳動物細胞が認識する転写終止およびポリアデニル化配列は、翻訳停止コドンの3'に位置する調節領域であり、よってプロモーター要素と一諸にコード配列に隣接している。成熟mRNAの3'末端は位置特異的翻訳後開裂およびポリアデニル化によって形成される。転写ターミネーターおよびポリアデニル化シグナルの例はSV40から誘導したものを包含する。

20

【0223】

哺乳動物宿主およびその他の宿主に外因性核酸を導入する方法は当分野で周知であり、使用する宿主細胞によって変わる。技術には、デキストラン仲介トランスフェクション、燐酸カルシウム沈殿、ポリブレン仲介トランスフェクション、プロトプラスト融合、電気穿孔、ウイルス感染、リボソームへのポリヌクレオチドのカプセル化、および核内へのDNAの直接的マイクロインジェクションがある。

【0224】

好ましい態様では、候補変異体タンパク質または候補変異体ライブラリータンパク質を細菌系で発現させる。細菌発現系は当分野で周知である。

【0225】

好適な細菌プロモーターは、細菌RNAポリメラーゼと結合でき、ライブラリータンパク質のコード配列からmRNAへの下流(3')方向への転写を開始できる任意の核酸配列である。細菌プロモーターは、通常コード配列の5'末端の近位に位置する転写開始領域を持っている。この転写開始領域は典型的にはRNAポリメラーゼ結合部位と転写開始部位を包含する。代謝経路の酵素をコードしている配列は特に有用なプロモーター配列を提供する。例には、ガラクトース、乳糖およびマルトースといった糖を代謝する酵素から誘導されるプロモーター配列、ならびにトリプトファンのような生合成酵素から誘導される配列がある。バクテリオファージ由来のプロモーターも使用でき、当分野で知られている。加えて、合成プロモーターおよびハイブリッドプロモーターもまた有用であり、例えばtacプロモーターはtrpおよびlacプロモーター配列のハイブリッドである。さらに、細菌プロモーターは、細菌RNAポリメラーゼと結合し転写を開始させる能力を持つ、非細菌由来の天然に存在するプロモーターを包含できる。

30

40

機能的プロモーター配列に加えて、有効なリボソーム結合部位が望ましい。E.coliでは、リボソーム結合部位をShine-Delgarno (SD) 配列と呼び、開始コドンとこの開始コドンの3-11ヌクレオチド上流に位置する3-9ヌクレオチド長の配列を包含する。

【0226】

発現ベクターはさらに、細菌中でライブラリータンパク質の選択を可能にするシグナルペプチド配列を包含できる。当分野で周知のように、シグナル配列は典型的にはその細胞からのタンパク質の分泌を指令する疎水性アミノ酸から成るシグナルペプチドをコードしている。タンパク質は増殖培地中に分泌される(グラム陽性菌)か、または細胞の内膜と外

50

膜の間にある周辺腔に分泌される（グラム陰性菌）。

【0227】

細菌発現ベクターはさらに、形質転換された細菌菌株の選択を可能にする選択マーカー遺伝子をも含む。好適な選択遺伝子は、アンピシリン、クロラムフェニコール、エリスロマイシン、カナマイシン、ネオマイシンおよびテトラサイクリンといった薬物に対して細菌を耐性とする遺伝子を包含する。選択マーカーはまた、生合成遺伝子、例えばヒスチジン、トリプトファンおよびロイシン生合成系路の遺伝子をも包含する。

【0228】

これらの成分を構築して発現ベクターとする。細菌用の発現ベクターは当分野で周知であり、とりわけ*Bacillus subtilis*、*E. coli*、*Streptococcus cremoris*、および*Streptococcus lividans*のためのベクターを包含する。 10

【0229】

細菌発現ベクターは、当分野で周知の技術、例えば塩化カルシウム処理、電気穿孔などを用いて細菌宿主細胞中に導入する。

【0230】

或る態様では、候補変異体タンパク質を昆虫細胞において生成させる。昆虫細胞の形質転換のための発現ベクター、特にバキュロウイルス系発現ベクターは当分野でよく知られており、例えばO'Reilly et al., *Baculovirus Expression Vectors: A Laboratory Manual* (New York: Oxford University Press, 1994)に記載されている。

【0231】

好ましい態様では、候補変異体タンパク質を酵母細胞において生成させる。酵母発現系は当分野で周知であり、*Saccharomyces cerevisiae*、*Candida albicans*および*C. maltosa*、*Hansenula polymorpha*、*Kluyveromyces fragilis*および*K. lactis*、*Pichia guilliermondii*および*P. pastoris*、*Schizosaccharomyces pombe*、ならびに*Yarrowia lipolytica*のための発現ベクターを包含する。酵母での発現のための好ましいプロモーター配列は、GAL1, 10誘導プロモーター、アルコールデヒドロゲナーゼ、エノラーゼ、グルコキナーゼ、グルコース-6-リン酸イソメラーゼ、グリセルアルデヒド-3-リン酸デヒドロゲナーゼ、ヘキソキナーゼ、ホスホフルクトキナーゼ、3-ホスホグリセラートムターゼ、ピルビン酸キナーゼ、および酸ホスファターゼ遺伝子由来のプロモーターを包含する。酵母選択マーカーは、ツニカマイシンに対する耐性を付与するADE2、HIS4、LEU2、TRP1、およびALG7；G418に対する 30耐性を付与するネオマイシンホスホトランスフェラーゼ遺伝子；および、酵母を銅イオン存在下で増殖可能とするCUP1遺伝子を包含する。

【0232】

好ましい態様では、候補変異体タンパク質または候補変異体ライブラリータンパク質を植物細胞中で発現させる。トランスジェニック植物での発現を意図した遺伝子配列を、まず、植物中で発現可能な適当なプロモーターに隣接する発現カセットに組み込む。この発現カセットはトランスジーンの発現のために必要なまたは選択されたさらなる配列を含むこともできる。このような配列には、転写ターミネーター、発現を増強するイントロンのような外来配列、エンハンサー配列、および、遺伝子産物が特定の細胞小器官および細胞コンパートメントを標的とすることを意図した配列を包含するが、これらに限定される訳ではない。次いでこれらの発現カセットを下記のような植物形質転換ベクターへと容易に移行させることができる。以下は、典型的な発現カセットの様々な構成成分の記載である。 40

【0233】

発現カセットに使用するプロモーターの選択は、トランスジェニック植物中のトランスジーンの空間的および時間的発現パターンを決定する。選択されたプロモーターはトランスジーンを特定の細胞型（例えば葉の表皮細胞、葉肉細胞、根の皮層細胞）または特定の組織もしくは器官（例えば根、葉または花）で発現させ、故にプロモーターの選択は遺伝子産物の蓄積を望む位置に基づくものである。本発明の好ましい態様では、オレオシン-TR融合タンパク質、オレオシン-TR融合タンパク質またはオレオシン-ハイブリッドTR/TRレダクターゼ融合タンパク質の発現に種子特異的プロモーターを使用する。最も好ましい態 50

様では、この種子特異的プロモーターはファセオリンプロモーターである。

【0234】

プロモーターは転写を促進する能力に違いがある。利用する宿主細胞に応じて、当分野で知られる幾つかの適当なプロモーターのいずれかを使用できる。構成的発現のためには、CaMV 35Sプロモーター、米アクチンプロモーター、またはユビキチンプロモーターが使用できる。これに代わり、誘導的プロモーターを選択して様々な誘導条件の下で遺伝子発現をさせることができる。化学誘導的発現のためにはタバコまたはArabidopsis由来の誘導的PR-1プロモーターが使用できる（例えば米国特許No. 5,689,044を参照されたい）。

【0235】

様々な転写ターミネーターが、核遺伝子発現カセットでの使用のために利用でき、トランスジーンを超えた転写の停止およびその正しいポリアダニル化を担っている。適当な転写ターミネーターとは、植物中で機能することが分かっているものであり、CaMV 35Sターミネーター、tm/ターミネーター、ノパリンシンターゼ（nos）ターミネーターおよびエンドウ豆rbcS E9ターミネーターを包含する。これらは単子葉植物と双子葉植物の両方に使用できる。好ましい態様ではファセオリン転写ターミネーターを使用する。プラスチドでの発現は停止を必要としないことがあるが、翻訳開始、伸長およびRNA安定性のために正しい5'および3'シグナルを必要とすることがある。

【0236】

数多くの配列が、転写ユニット内部からの遺伝子発現を促進することが見出されており、これらの配列を本発明に係る遺伝子と共に使用して、トランスジェニック植物中での発現を増大させることができる。例えば、種々のイントロン配列、例えばトウモロコシAdhl遺伝子のイントロンは、特に単子葉植物細胞で発現を増強することが示されている。加えて、ウイルス由来の幾つかの非翻訳リーダー配列もまた発現を増強することが分かっており、これらは双子葉植物細胞で特に有効である。トランスジェニック植物でそれらを発現させるためには、使用するDNA分子のコード配列は、特にそのDNA分子が原核生物由来である場合、修飾および最適化が必要となるかも知れない。全ての生物はコドンの使用に特定の好みがあることが当分野で知られており、本発明に係るDNA分子のヌクレオチド配列中のコドンは、それによってコードされているアミノ酸を維持しつつ、特定の植物の好みに適合するように変えることができる。少なくとも35%、好ましくは45%以上のGC含有量を持つコード配列から、植物での高発現が最良に達成できる。低GC含有量のヌクレオチド配列は、メッセージを不安定化し得るATTTAモチーフ、そして不適切なポリアダニル化を惹起し得るAATAAAモチーフが存在するためうまく発現しない。好ましい遺伝子配列は単子葉植物と双子葉植物種の両方で充分発現されるが、配列は、特定のコドンの好みならびに単子葉植物または双子葉植物のGC含有量の傾向を考慮して修飾することができる。何故ならこれらの好みは相違することが示されているためである（Murray et al. (1989) Nucl Acids Res 17: 477-498）。加えて、メッセージの末端切除を惹起する異常スプライス部位の存在について、このヌクレオチド配列をスクリーニングする。上記のようなヌクレオチド配列内部でなされることが求められる変化の全ては、位置指定突然変異誘発、PCR、ならびに、例えば公開されている特許出願EP 0 385 962、EP 0 359 472、およびWO 93/07278（これらの開示の全体は引用によりここに含める）に記載の方法を用いた合成遺伝子構築といった周知の技術を用いて実施する。

【0237】

翻訳を有効に開始させるため、開始メチオニンに隣接する配列は修飾が必要であるかも知れない。例えば、これらは植物において有効であることが分かっている配列を含めることにより修飾できる。Joshiは植物のための適当なコンセンサスを示唆しており（Nuc Acids Res (1987) 15:6643-6653）、さらなるコンセンサス翻訳イニシエーター（Clontech 1993/1994 catalog, page 210）を含めることができる。これらのコンセンサス配列は本発明に係るヌクレオチド配列と共に使用するのに適している。この配列は、ATGまで（これを含む）（第二のアミノ酸は修飾しないでおく）、或いはATGの後のGTCまで（これを含む）（トランスジーン第二のアミノ酸を修飾する可能性がある）のヌクレオチド配列を含む構築

物中に組み込む。

【0238】

遺伝子産物をターゲッティングするための様々なメカニズムが植物に存在することが分かっており、これらのメカニズムの機能を調節する配列が、幾分詳細に特性決定されている。例えば、葉緑体への遺伝子産物ターゲッティングは、葉緑体へと運ばれる間に開裂して成熟タンパク質を生成する、種々のタンパク質のアミノ末端に見出される一時的配列によって調節される (Comai et al. (1988) J Biol Chem 263: 15104-15109)。別の遺伝子産物は、ミトコンドリアやペルオキシソームのような他の細胞小器官に局在する (Unger et al. (1989) Plant Mol Biol 13:411-418)。これらの生成物をコードしているcDNAを操作して、ヘテロローガス遺伝子産物をこれらの細胞小器官へとターゲッティングすることができ、加えて、遺伝子産物を他の細胞コンパートメントにターゲッティングする配列が特性決定されている。

10

【0239】

アミノ末端配列は、ER、アポプラストへのターゲッティング、およびアリュールン細胞からの細胞外分泌を担っている (Koehler & Ho (1990) Plant Cell 2:769-783)。さらに、カルボキシ末端配列と共にアミノ末端配列は、遺伝子産物の液胞ターゲッティングを担っている (Shinshi et al., (1990) Plant Mol Biol 14:357-368)。上に述べた適当なターゲッティング配列を目的のトランスジーン配列と融合させることにより、トランスジーン生成物を所望の細胞小器官または細胞コンパートメントへと向かわせることが可能である。

20

【0240】

別の好ましい態様では、本発明に係るDNA分子を直接プラスチドゲノム中に導入する。プラスチド形質転換技術は米国特許Nos. 5,451,513、5,545,817、5,545,818および5,576,198; PCT出願nos. WO 95/16783およびWO 97/32977; ならびにMcBride et al., Proc Natl Acad Sci USA 91: 7301-7305 (1994)(これらの全開示は引用によりここに含める) に詳細に記載されている。或る態様では、プラスチド形質転換は、バイオリスティクスで達成し; まず単細胞緑藻類 *Chlamydomonas reinhardtii* で実施し (Boynton et al. (1988) Science 240:1534-1537) 次に *Nicotiana tabacum* に拡大し (Svab et al. (1990) Proc Natl Acad Sci USA 87:8526-8530)、シス作動性抗生物質耐性遺伝子座についての選択 (スペクチノマイシンまたはストレプトマイシン耐性) または非光合成突然変異表現型の補完につ

30

【0241】

別の態様では、抗生物質耐性選択マーカーに隣接するクローンプラスチドDNAを用いて、葉もしくはカルス組織の粒子衝突、またはプロトプラストによるプラスミドDNAのポリエチレングリコール (PEG) 仲介取り込みによって、タバコプラスチド形質転換を実施する。ターゲッティング配列と命名された1ないし1.5kbフランキング領域は、プラスチドゲノムとの相同的組換えを促進し、156kbタバコプラスチドゲノムの特異的領域の置換または修飾を可能にする。最初に、スペクチノマイシンおよび/またはストレプトマイシンに対する耐性を付与するプラスチド16S rDNAおよびrps12遺伝子における点突然変異を、形質転換のための選択マーカーとして利用し (Svab et al. (1990) Proc Natl Acad Sci USA 87:8526-8530; Staub et al. (1992) Plant Cell 4:39-45、これらの開示全体を引用によりここに含める)、およそ100回の標的葉の衝突につき1回の頻度で安定なホモプラスミク形質転換体を得た。これらのマーカー間のクローニング部位の存在は、外来遺伝子の導入のためのプラスチドターゲッティングベクターの創出を可能にする (Staub et al. (1993) EMBO J 12:601-606、この開示全体を引用によりここに含める)。劣性rRNAまたはrタンパク質抗生物質耐性遺伝子を優性選択マーカー、スペクチノマイシン無毒化酵素アミノグリコシド-3'-アデニルトランスフェラーゼをコードしている細菌aadA遺伝子で置換することにより、形質転換頻度の実質的増大を達成した (Svab et al. (1993) Proc Natl Acad Sci USA 90: 913-917、この開示全体を引用によりここに含める)。かつてこのマーカーは、緑藻 *Chlamydomonas reinhardtii* のプラスチドゲノムの高頻度形質転換に成功裏に用

40

50

いられていた(Goldschmidt-Clermont, M. (1991) Nucl Acids Res 19, 4083-4089、この開示全体を引用によりここに含める)。近年、タバコおよびコケ類Physcomitrella由来のプロトプラストのプラスチド形質転換が、PEG伸介DNA取り込みを用いて達成されている(0'Neill et al. (1993) Plant J 3:729-738; Koop et al. (1996) Planta 199:193-201、これらの開示全体を引用によりここに含める)。

【0242】

粒子衝突とプロトプラスト形質転換は共に本発明の状況において適当である。油料種子植物のプラスチド形質転換がArabidopsisおよびBrassica属で成功裏に実施された(Sikdar et al. (1998) Plant Cell Rep 18:20-24; PCT出願W0 00/39313、これらの全開示を引用によりここに含める)。

10

【0243】

本発明に係るDNA分子を、植物プラスチド中でこのDNA分子を発現できるプロモーターを含むプラスチド発現カセット中に挿入する。植物プラスチド中での発現が可能な好ましいプロモーターは、例えばプラスチド遺伝子のコード領域の上流の5'フランキンク領域から単離したプロモーターであり、これは同じまたは異なる種から誘導したものであってよく、その天然の産物は、非緑色組織に存在するものを包含する大多数の種類のプラスチドに典型的に見出される。プラスチドにおける遺伝子発現は核遺伝子発現とは異なり、原核生物における遺伝子発現に関連している(Stern et al. (1997) Trends in Plant Sci 2:308-315、この全開示を引用によりここに含める)。

【0244】

プラスチドプロモーターは一般に、原核生物プロモーターに典型的な-35および-10要素を含み、PEP(プラスチドコード化RNAポリメラーゼ)プロモーターと呼ばれる幾つかのプラスチドプロモーターは、大抵プラスチドゲノム中でコードされているE.coli様RNAポリメラーゼによって認識され、一方、NEPプロモーターと呼ばれる別のプラスチドプロモーターは核コード化RNAポリメラーゼによって認識される。いずれのタイプのプラスチドプロモーターも本発明にとって好適である。プラスチドプロモーターの例は、clpP遺伝子のプロモーター、例えばタバコclpP遺伝子プロモーター(W0 97/06250、この全開示を引用によりここに含める)およびArabidopsis clpP遺伝子プロモーター(米国出願No. 09/038,878、この全開示を引用によりここに含める)を包含する。植物プラスチドにおいてDNA分子の発現を可能にする別のプロモーターは、プラスチド16SリボソームRNAオペロンの調節領域由来のものである(Harris et al., (1994) Microbiol Rev 58:700-754; Shinozaki et al. (1986) EMBO J 5:2043-2049、これらの全開示を引用によりここに含める)。植物プラスチドにおいてDNA分子の発現を可能にするプロモーターの別の例は、psbAプロモーターまたはam rbcLプロモーターを包含する。プラスチド発現カセットは、好ましくは、本発明に係るDNA分子と機能的に連結したプラスチド遺伝子3'非翻訳配列(3' UTR)をさらに含む。非翻訳配列の役割は、好ましくは転写の終止ではなく転写されたRNAの3'プロセッシングを指令することである。好ましくは、この3' UTRはプラスチドrps16遺伝子3'非翻訳配列、またはArabidopsisプラスチドpsbA遺伝子3'非翻訳配列である。さらなる好ましい態様では、プラスチド発現カセットは、3'非翻訳配列の代わりにポリG列を含む。プラスチド発現カセットはまた、好ましくは、本発明に係るDNA分子に機能的に連結した、植物プラスチド中で機能的な5'非翻訳配列(5' UTR)をさらに含む。

20

30

40

【0245】

プラスチド発現カセットはプラスチド形質転換ベクター中に含まれるが、これは好ましくは、相同的組換えによるプラスチドゲノム中への組み込みのためのフランキンク領域をさらに含む。プラスチド形質転換ベクターは所望により少なくとも1個のプラスチド複製起点を含む。本発明はさらに、該DNA分子が該植物プラスチド中で発現され得る、係るプラスチド形質転換ベクターによって形質転換された植物プラスチドを包含する。本発明はさらに、この植物プラスチドを含む植物または植物細胞(その子孫を包含する)を包含する。好ましい態様では、その子孫を包含する該植物または植物細胞は、トランスジェニックプラスチドについてホモプラスミックである。

50

【0246】

植物プラスチドにおいてDNA分子の発現をさせることができるその他のプロモーターには、好ましくは該植物または発現が行われる植物細胞の細胞下小器官または構成成分に関してヘテロローガスである、トランスアクチベーターの調節を受けるプロモーターがある。これらの場合、トランスアクチベーターをコードしているDNA分子を適当な核発現カセット中に挿入し、これを植物核DNAに導入する。このトランスアクチベーターはプラスチド一過性ペプチドを用いてプラスチドにターゲッティングする。トランスアクチベーターとトランスアクチベーターにより駆動するDNA分子は、選択したプラスチド形質転換された系統を、プラスチドターゲッティング配列が補完された核プロモーターに機能的に連結しているトランスアクチベーターをコードしているDNA分子を含むトランスジェニック系統と交雑することによって、または、所望DNA分子を含むプラスチド形質転換ベクターを、核プロモーターに機能的に連結したプラスチドターゲッティング配列を補完したトランスアクチベーターをコードしているDNA分子を含むトランスジェニック系統中に導入することによって、一諸にする。核プロモーターが誘導的プロモーター、とりわけ化学誘導的プロモーターである場合、植物プラスチド中でのDNA分子の発現は化学的インデューサーを葉に適用することで活性化される。このような誘導的トランスアクチベーター仲介プラスチド発現系は、しっかりと調節ができ、誘導前には検出可能な発現が無く、誘導後には極めて高いタンパク質の発現と蓄積のあることが好ましい。好ましいトランスアクチベーターは、例えばウイルスRNAポリメラーゼである。このタイプの好ましいプロモーターは、単一サブユニットRNAポリメラーゼにより認識されるプロモーター、例えばバクテリオファージT7 DNA依存RNAポリメラーゼにより認識されるT7遺伝子10プロモーターである。T7ポリメラーゼをコードしている遺伝子を好ましくは核ゲノム中に導入し、このT7ポリメラーゼをプラスチド一過性ペプチドを用いてプラスチドにターゲッティングする。遺伝子、例えばT7ポリメラーゼのようなウイルスRNAポリメラーゼをコードしている遺伝子の核発現に好適なプロモーターは上記および本出願の他所に記載がある。プラスチドにおけるDNA分子の発現は構成的であっても誘導的であってもよく、係るプラスチド発現もまた器官または組織特異的であってもよい。種々の発現系の例はW0 98/11235に詳細に記載されており、この全開示を引用によりここに含める。したがって或る側面では、本発明は、例えば米国特許No. 5,614,395（この全開示を引用によりここに含める）に記載のように、T7遺伝子10プロモーター/ターミネーター配列により調節される葉緑体リポータートランスジェニックと機能的に連結した、化学誘導的PR-1aプロモーター、例えばタバコのPR-1プロモーターの調節下に、葉緑体ターゲッティングファージt7 RNAポリメラーゼの核ゲノムにおける共役発現を利用した。別の態様では、母系遺伝したTR遺伝子についてホモプラスミックなプラスチド形質転換体を、核でT7ポリメラーゼを発現する系統により授粉させる時、両方のトランスジェニック構築物を持つが、PR-1aインデューサー化合物であるベンゾ(1,2,3)チアジアゾール-7-カルボチオ酸S-メチルエステル（BTH）の葉への適用により、プラスチドにおける酵素的に活性なタンパク質の大量合成が誘発されるまではそれが発現されないF1植物が得られる。

【0247】

好ましい態様では、2またはそれ以上の遺伝子、例えばTR遺伝子をオペロン様ポリシストロン遺伝子中の1個のプロモーター由来のプラスチドゲノムから転写する。好ましい態様では、このオペロン様ポリシストロン遺伝子は、オペロン様ポリシストロン遺伝子中の2個の遺伝子の間に介在DNA配列を含んでいる。好ましい態様では、このDNA配列は、プラスチド配列との相同的組換えを回避するため、プラスチドゲノムには存在しない。別の好ましい態様では、このDNA配列を、非真核生物遺伝子の5'非翻訳（UTR）領域から、好ましくはウイルス5'UTRから、好ましくはバクテリオファージ、例えばT7、T3またはSP6ファージから誘導した5'UTRから誘導する。好ましい態様では、オペロン様ポリシストロン遺伝子のRNA転写物において、例えば該DNA配列と下流遺伝子のRBSの間にRNA二次構造の形成を妨げるために、このDNA配列の一部を修飾できる。このような二次構造は下流遺伝子の発現、とりわけ翻訳の開始を阻害または抑制できる。このようなRNA二次構造は、「mfold」プ

10

20

30

40

50

ログラムバージョン3 (Zuker and Turner, Washington University School of Medicine, St-Louis, MOより入手できる) のようなコンピューターモデルおよびプログラムならびに当業者の知悉するその他の方法を用いて融点を測定することにより予想する。

【0248】

オペロン様ポリシストロン遺伝子中の介在DNA配列の存在は下流遺伝子のRBSの利用可能性を増大させ、したがって高い発現速度をもたらす。このような戦略は、オペロン様キメラ遺伝子中の単一プロモーター由来のプラスチドゲノムから転写される任意の2またはそれ以上の遺伝子に適用できる。

【0249】

植物の形質転換に利用できる形質転換ベクターが当業者に多数知られており、本発明に関連する遺伝子をこのような任意のベクターと共に使用することができる。ベクターの選択は好ましい形質転換技術および形質転換される標的の種に依存する。或る標的の種については異なる抗生物質または除草剤選択マーカーが好ましいかも知れない。

【0250】

形質転換で常套的に用いられる選択マーカーは、カナマイシンおよび関連抗生物質に対する耐性を付与するnptII遺伝子 (Messing & Vieira. (1982) Gene 19:259-268; Bevan et al. (1983) Nature 304:184-187)、除草剤ホスフィノスリシンに対する耐性を付与するbar遺伝子 (White et al. (1990) Nucl Acids Res 18: 1062; Spencer et al. (1990) Theor Appl Genet 79:625-631)、抗生物質ハイグロマイシンに対する耐性を付与するhph遺伝子 (Yanofsky, et al. (1992) Gene 117:161-167)、メソトレキサートに対する耐性を付与するdhfr遺伝子 (Bourouis et al., EMBO J. 7:1099-1104 (1983))、グリホサートに対する耐性を付与するEPSPS遺伝子 (米国特許Nos. 4,940,935および5,188,642)、および通常は植物毒性である糖マンノースに対する耐性を付与するマンノースホスファートイソメラーゼ遺伝子pmi (Negrotto, et al. (2000) Plant Cell Rep 19:798-803) を包含する。

【0251】

*Agrobacterium tumefaciens*を使用する形質転換には多くのベクターが適している。これらは典型的には少なくとも1個のT-DNA境界配列を持ち、pBIN19 (Bevan, (1984) Nucl Acids Res)およびpXYZのようなベクターを包含する。*Agrobacterium*の形質転換に好適な典型的ベクターは、バイナリーベクターpCIB200およびpCIB2001、ならびにバイナリーベクターpCIB10およびそのハイグロマイシン選択誘導体である (米国特許No. 5,639,949)。

【0252】

*Agrobacterium tumefaciens*を使用しない形質転換は、選択された形質転換ベクターにおけるT-DNA配列の必要性を回避する。その結果、これらの配列を欠くベクターを、上記のT-DNA含有ベクターのようなベクターの代替として使用できる。*Agrobacterium*に依拠しない形質転換技術には、粒子衝突、プロトプラスト取り込み、例えばPEGおよび/または電気穿孔、ならびにマイクロインジェクションがある。ベクターの選択は、形質転換を受ける好ましい種の選択にほぼ依存する。非*Agrobacterium*形質転換に好適な典型的ベクターはpCIB3064、pSOG19、およびpSOG35を包含する (米国特許No. 5,639,949)。

【0253】

目的のコード配列を発現系にクローニングしたならば、これを植物細胞に導入する。植物の形質転換と再生の方法は当分野で周知である。例えばTiプラスミドベクターが外来DNAのデリバリーに利用されており、DNAの直接取り込み、リポソーム、電気穿孔、マイクロインジェクション、およびマイクロプロジェクトイルもまた利用されている。さらに、*Agrobacterium*属由来の細菌を植物細胞の形質転換に利用することもできる。

【0254】

双子葉植物の形質転換技術は当分野で周知であり、*Agrobacterium*に基づく技術と*Agrobacterium*を必要としない技術を包含する。非*Agrobacterium*技術は、プロトプラストまたは細胞による外因性遺伝物質の直接的取り込みを含む。これはPEGまたは電気穿孔仲介取り込み、粒子衝突仲介デリバリー、またはマイクロインジェクションによって達成する。いずれの場合にも、形質転換を受けた細胞を当分野で既知の標準技術を用いて全植物体に再

生する。

【0255】

多くの双子葉植物および単子葉植物種を形質転換する方法が当分野でよく知られている。好ましい技術は、PEGまたは電気穿孔技術を用いるプロトプラスト中への直接的遺伝子移動、カルス組織中への粒子衝突、ならびにAgrobacterium仲介形質転換を包含する。

【0256】

さらに、当分野で周知の技術を用いて候補変異体ライブラリータンパク質を融合タンパク質として製造することもできる。例えば、変異体タンパク質を該タンパク質の発現を増大させまたは安定化する他のタンパク質に融合させることができる。同様に、他の融合相手、例えば抗体、ライブラリー成員を細胞の細胞下または細胞外コンパートメントに局在させるターゲッティング配列、ライブラリータンパク質またはそれをコードしている核酸のいずれかを精製または単離可能とする救出配列もしくは精製タグ、安定性または分解からの防御を付与する安定性配列、リポーター、検出および選択遺伝子もしくはタンパク質を包含する融合タンパク質、またはこれらの組み合わせ、ならびに必要な応じてリンカー配列が使用できる。

10

【0257】

好ましい態様では、候補変異体タンパク質または候補変異体ライブラリータンパク質を発現後に精製または単離する。試料中にどのような他の成分が存在するかに応じて、当業者の知悉する様々な方法で変異体タンパク質を単離精製できる。標準的精製法は、電気泳動、分子、免疫およびクロマトグラフィー技術（イオン交換、疎水性、アフィニティー、および逆相HPLCクロマトグラフィーを包含する）、ならびにクロマトフォーカシングを包含する。タンパク質濃度に関連した限外濾過およびダイアフィルトレーション技術もまた有用である。適当な精製技術の一般的指針については、Scopes, R., Protein Purification, Springer-Verlan, NY (1982) を参照されたい。必要な精製の程度はその変異体タンパク質の用途に依存する。幾つかの例では精製は必要ない。

20

【0258】

製造したならば、この変異体TRタンパク質は実験的に試験しインビボおよびインビトロ検定で評価できる。適当な検定は、一次および二次スクリーニング検定ならびに精製タンパク質動態パラメータ、即ち K_{cat} および K_m の特性決定を包含する。(図11および12を参照されたい)。

30

【0259】

製造したならば、この本発明に係る変異体TRタンパク質および核酸は幾つかの適用に用途が見出せる。好ましい態様では、この変異体TRを使用して、小麦、ライ麦および大麦中のグルテンの抗原性を低下させる。

【0260】

別の態様では、変異体TRを使用して、蛇毒、ミツバチ、サソリおよび細菌神経毒破傷風およびボツリヌス中毒に見られる毒性タンパク質のジスルフィド結合を還元する。

【0261】

好ましい態様では、変異体TRを使用して代替基質を還元する。チオレドキシンレダクターゼのための有用な代替基質は、チオレドキシンドメインを含むことが判明した幾つかの植物および哺乳動物タンパク質を包含する。例えば、タンパク質ジスルフィドイソメラーゼ(PDI)はチオレドキシンに相同的な内部配列を示す2個の領域を含んでいる。PDIはチオレドキシンレダクターゼの基質である。タンパク質ジスルフィドイソメラーゼは、ウシ(Yamauchi et al., Biochem. Biophys. Res. Commun. 146:1485-1492, 1987)、ニワトリ(Parkkonen et al., Biochem. Zn 256:1005-1011, 1988)、ヒト(Rapilajaniemi et al. EMB J. 6:643-649 1987)、マウス(Gong, et al., Nucleic Acids Res. 16:1203, 1988)、ウサギ(Fliegel et al., J. Biol. Chem. 265:15496-15502, 1990)、およびラット(Edman et al., Nature 317:267-270, 1985)といった哺乳動物供給源から同定されている。PDIは酵母から単離された(Tachikawa et al., J. Biochem. 110:306-313)。適当なPDIはW09501425公開19950112およびW09500636公開19950105、ならびにヒトおよび植物型を包

40

50

含する当分野で既知のその他のPDIに見出すことができる。

【0262】

チオレドキシンレダクターゼの基質である酸化還元剤、例えばチオレドキシンおよびPDIの組成物および用途は当分野で既知であり、本明細書に記載してある。ジスルフィド結合は、酵素、構造タンパク質などといった多くの種類のタンパク質に存在している。酵素とはプロテアーゼ、アミラーゼなどといった触媒タンパク質であり、一方構造タンパク質とはケラチンなどのような硬タンパク質であってよい。体毛、羊毛、皮膚、皮革、獣皮、食品、飼料中のタンパク質物質は染料であり、ヒト組織はジスルフィド結合を含む。PDIおよびチオレドキシン、および酸化還元相手による、これらの物質のうち幾つかの処理が過去に記載されている。例として、人間および動物の体毛にウェーブを付け、まっすぐにし、除去しそして柔らかにするためのチオレドキシンの使用がEP183506およびW08906122に記載されている。米国特許4771036はまた、白内障の防止および逆転のためのチオレドキシンの使用を記載している。生体反応において金属で触媒される酸化的損傷を防止するためのチオレドキシンの使用がPigiet et al. EP 237189に記載されている。EP272781およびEP276547はそれぞれ人間の体毛の配置変更、および羊毛の処理のためのPDIの使用を記載している。このような酵素の使用は全て、遊離のタンパク質スルヒドリル基へのタンパク質ジスルフィド結合の還元および/または同一ポリペプチド内でのまたは異なるポリペプチド間でのジスルフィド結合の再配置と関連している。したがって、本発明に係るチオレドキシンレダクターゼを、所望によりその補助因子NADHまたはNADPHと共に酸化還元相手としての係る組成物に添加し、その酸化還元剤を再生し、そのようにしてその組成物の有用性を向上させることができる。これに代わる態様では、本発明に係るチオレドキシン変異体を、本明細書に教示する酸化還元剤とのタンパク質融合物として提供する。例えばこの組成物は、硬タンパク質、特に体毛、皮膚および羊毛の処理または分解のために、獣皮の除毛および軟化のために、洗剤添加物として織物の処理およびクリーニングのために、食品および飼料の濃厚化およびゲル化のために、パンまたはペストリー製品中のグルテンの強化のために、そして眼の苦痛を軽減する医薬として使用できる。本発明に係る組成物、特にPDIとの組成物を、分子間タンパク質ジスルフィド架橋を生成する物質を含有する他のタンパク質と共に使用して、高分子量またはゲル化した組成物を得ることができる。したがって本発明は、生魚肉ペースト、蒲鉾（フィッシュケーキ）、魚肉/畜肉ソーセージ、豆腐（大豆カード）、麺類、菓子類、パン、練り生地、食品接着剤、シート様肉食品、ヨーグルト、ゼリーおよびチーズといった食品加工の分野で使用できる。加えてこれらは、化粧品、マイクロカプセルの素材および固定化酵素の担体を包含する広範囲の工業における新規なタンパク質由来材料として使用できる。

10

20

30

40

【0263】

好ましい態様では、変異体TR-オレオシン-チオレドキシンおよびオレオシン-変異体チオレドキシンレダクターゼ融合タンパク質は、油体と共に蓄積する。別の態様では、オレオシン-チオレドキシン/変異体チオレドキシンレダクターゼハイブリッド融合タンパク質は、油体と共に蓄積する。この油体を分画して融合タンパク質の部分精製を達成できる。融合タンパク質の付随する精製された油体は、チオレドキシンおよびチオレドキシン-レダクターゼ活性ならびに皮膚（化粧品）または食品用途の適用における機能的利益を試験するための成分として使用できる。油体は化粧品および食品成分にとって極めて好適な加工および調合性質を持っている。故に、油体と結びついたオレオシン融合物としてチオレドキシンおよび/またはチオレドキシンレダクターゼをデリバリーすることは、加工を単純にし、そして製品の安定性を増大させる。

【0264】

別の態様では、油体からチオレドキシンまたはチオレドキシンレダクターゼを精製するため、第二の精製工程を実施できる。この事は、チオレドキシンおよびチオレドキシンレダクターゼの活性を試験するための、そして化粧品または食品用途における機能的利益を提供するための成分として使用できる、該タンパク質の高度精製調製品を導く。米国特許公開No. 2002/0037303をも参照されたい（引用によりここに含める）。

50

【0265】

本明細書に論ずるその他の調合物および組成物、例えば油体の態様に加えて、本発明に係る組成物は可溶性チオレドキシンレダクターゼおよび/または酸化還元剤、ならびに当分野で既知のその他の成分、例えば賦形剤、安定剤、増量剤、洗浄剤などを含有できる。この組成物は任意の簡便な形態、例えば粉末、ペースト、液体または顆粒形態として調合できる。この酵素は酵素安定剤を含有させることによって液体で安定化させることができる。通常、この組成物の溶液のpHは5-10、幾つかの例では7.0-8.5である。用途に応じてしばしば無菌組成物が好ましい。

【0266】

さらに、家畜飼料中の穀物および穀物由来製品の品質もまた、分子間および分子内ジスルフィド結合による影響を受ける。穀物の消化性、栄養利用性、および抗栄養因子（例えばプロテアーゼ、アルニラーゼインヒビターなど）の中和は、ジスルフィド結合の程度を低下させることによって増大するであろう。(1999年12月15日出願のW0 00/36126を参照されたい)。トウモロコシおよび大豆における、所望によりチオレドキシンを伴うトランスジェニックチオレドキシンレダクターゼ変異体の発現ならびに穀物加工、例えば湿潤製粉の際のチオレドキシンレダクターゼの使用は、工業的加工中またはこれに先立って種子タンパク質中のジスルフィド結合を還元する代替法を提供する。故に本発明は、貯蔵タンパク質の品質が変化した穀物、および、工業的加工の際または動物が消化する際（いずれも以降「加工」と称する）通常穀物とは質的に異なる挙動をする穀物を提供する。所望によりチオレドキシンを伴うこのチオレドキシンレダクターゼのデリバリー法は、加工中に添加するための外因性チオレドキシンおよび/またはチオレドキシンレダクターゼ供給源を創出する必要性を排除する。穀物を介してチオレドキシンおよび/またはチオレドキシンレダクターゼを供給する第二の利点は、加工前にまたは余分な加工工程として、当該種子の貯蔵またはマトリックスタンパク質と該酵素を接触させるために種子の完全性を物理的に破壊する必要がないことである。本明細書に記載する発明は、全ての穀物収穫物、特にトウモロコシ、大豆、小麦、および大麦、最も具体的にはトウモロコシおよび大豆、とりわけトウモロコシに適用できる。所望によりチオレドキシンを伴うトランスジェニックチオレドキシンレダクターゼの穀物における発現は、穀物加工に移行する材料（種子）の品質を変え、穀物加工から誘導される材料の品質を変え、加工中の特定種子成分の収量を最大化し（効率を増大させ）、加工方法を変え、そして製粉の流れから種子由来部分または成分の新たな用途を作り出す手段である。したがって本発明は、好ましくは誘導性プロモーターの調節下に、例えば誘導性プロモーターと機能的に連結した、またはトランスアクチベーターにより調節されるプロモーターの調節下に（ここで、対応するトランスアクチベーターは該誘導性プロモーターの調節下にあるか、または第二の植物中で発現され該プロモーターがこの第二植物とのハイブリダイゼーションにより活性化される）、所望によりチオレドキシンを伴うチオレドキシンレダクターゼ変異体を発現する植物（ここでTRは好ましくは熱安定性でありまたは真核生物レダクターゼであり；係る植物はその種子をも包含し、この種子は所望により処理（例えば下処理または被覆されている）および/または包装され、例えば使用説明書と共に袋に入れられている）、および、そこから収穫した、例えば上記の製粉工程で使用するための種子を提供するものである。本発明に係るトランスジェニック植物は所望によりNADPHまたはNADHの生産を増強する遺伝子をさらに含む。

【0267】

さらに本発明は、上記の植物から収穫した種子から澱粉またはタンパク質を抽出することを含む、澱粉および/またはタンパク質を生成するための方法；および上記のチオレドキシンレダクターゼ発現植物由来の種子を浸漬し、澱粉および/またはタンパク質をそこから抽出することを含む、湿潤製粉のための方法を提供する。例えば好熱性生物、例えばarchaea由来の、例えばMethanococcus jannaschiiまたは本明細書に記載のArchaeoglobus fulgidus由来の熱安定性酵素が好ましい。穀物における、所望によりチオレドキシンを伴うトランスジェニックチオレドキシンレダクターゼ変異体の発現はまた、特に動物飼料における消化性に関連する穀物の性質の改善に有用である。プロテアーゼに対する飼料タンパク

質の感受性は時間とタンパク質コンホメーションの関数である。飼料調合物にはしばしば穀粒の挽き割りが用いられ、蒸気による薄片化もまた使用される。これらのプロセスはいずれも穀粒の消化性を助ける事を目的としている。その完全性が動物の胃内でより容易に破壊され得る軟穀粒が望ましい。コンホメーション制約およびタンパク質間の架橋がプロテアーゼ感受性の主たる決定要因である。増大させたチオレドキシンおよび/またはチオレドキシンレダクターゼ発現によりこれらの結合を修飾改変することは消化を助ける。タンパク質の含有量と質は薄片化粗粉の生産とmasaの生産の重要な決定要因である。ジスルフィド結合の還元は、小麦代用品、特に高タンパク質白色トウモロコシ変種から製造した粉としての使用に好適なようにトウモロコシ粉の性質を変化させる。米国の大豆収穫量の半分以上が粉碎または製粉されており、得られる低脂肪大豆粉または脱脂大豆粉（または大豆ミール）中のタンパク質の質はその後の加工にとって重要である。大豆加工の流れからのタンパク質の収量と質は経済的に重要であり、主としてタンパク質のコンホメーションに依存している。トランスジェニックチオレドキシンおよび/またはチオレドキシンレダクターゼの発現によるチオレドキシン活性の増大はタンパク質の溶解度を増大させ、よって水溶性タンパク質画分で収量を増加させる。回収は塩基性条件下での脱脂大豆ミールの水性抽出によって促進する。トランスジェニックチオレドキシンおよび/またはチオレドキシンレダクターゼの発現によるチオレドキシン活性の増強はまた、効率的抽出に必要なpHを低下させ、それにより水酸化カルシウムまたは水酸化ナトリウム投入量を減らし、後の酸沈殿化のための酸投入量をも低下させ、アルカリ損傷のない効率的なタンパク質回収を可能にし、水消費とプラント廃水（これは実質的な生物学的酸素要求量の負荷を含む）処理を減らす。タンパク質の酸化還元状態は、大豆タンパク質が供給する重要な機能的性質、例えば溶解度、吸水性、粘度、粘着/接着、ゲル化および弾性に影響を及ぼす。大豆タンパク質濃縮液の生産と、大豆タンパク質単離物のプロテアーゼによる加水分解の間の繊維除去は、本明細書に記載のチオレドキシン活性の増大によって向上する。同様に、トウモロコシについて上に述べたように、トランスジェニックチオレドキシンおよび/またはチオレドキシンレダクターゼの発現によるチオレドキシン活性の増大は、酵素活性大豆粉の機能性と、動物飼料中の大豆ミール画分および蒸気薄片化画分の消化性を向上させる。種子の発生の間の、そして加工の間のタンパク質品質の修飾もまた提供されるが、種子発生中のチオレドキシン活性の結果遭遇する可能性のある貯蔵タンパク質蓄積に及ぼす著しい副作用を回避するため、このトランスジェニックチオレドキシンおよび/またはチオレドキシンレダクターゼは細胞コンパートメントにターゲッティングされ、且つ上記のように熱安定性であることが好ましい。これとは別にこのチオレドキシンレダクターゼ変異体、および所望によりチオレドキシンは、加工用酵素として（または本明細書で教示する融合物として）添加することができるが、それは、（トウモロコシ湿潤製粉とは対照的に）ジスルフィド結合の破壊は穀粒の完全性破壊（粉碎および油抽出）後まで必要ないためである。上に述べたようにタンパク質ジスルフィドイソメラーゼ（PDI）もまたチオレドキシンにとって有用である。

【0268】

TRと共に油体を使用することに関して、「Thioredoxin and thioredoxin reductase containing oil body based products」なる標題のUS20020037303（公開20020328）を引用してここに含める。

【0269】

種子と増加のための本発明に係る酵素のさらなる用途をW00058453（公開20001005）に見出すことができる。チオレドキシンレダクターゼ変異体は、種子と穀物の品質向上のための本明細書に記載の用途のため、所望によりチオレドキシンと共に発現され、または外部から添加することができる。興味深いトランスジェニック植物は、大麦、小麦、Arabidopsis、タバコ、米、Brassica、Picea、または大豆、トウモロコシ、カラス麦、ライ麦、モロコシ、キビ、ライコムギ、ならびにまぐさおよび芝を包含する。本発明に係るトランスジェニック植物は、同種の非トランスジェニック植物の同部分に比して低下したアレルギー性を持ち得る。このアレルギー性は過敏性であってよく、この過敏性が少なくとも5%低

下している。さらに、本発明に係るトランスジェニック植物は、同種の非トランスジェニック植物の同部分に比して増大した消化性を持ち得る。この消化性は少なくとも5%増大している。トランスジェニック植物は、同種の非トランスジェニック植物の同部分に比してジベレリン酸誘導酵素の早期発現開始および/または増大した発現のある、該植物の少なくとも一部を持ち得る。好ましくは該酵素は、プルナーゼ、 α -アミラーゼである。この植物の一部は好ましくは可食部分、より好ましくは穀物または種子である。好ましいプロモーターは種子または穀物の成熟に特異的なプロモーターであり、例えば米グルテリン、米オリジン、米プロラミン、大麦ホルディン、小麦グリアジン、小麦グルテリン、トウモロコシゼイン、トウモロコシグルテリン、カラス麦グルテリン、モロコシカシリン、キビペニセチン、ライ麦セカリン、およびトウモロコシ胚特異的グロブリンより成る群から選ばれる。その他の態様には、本発明に係るトランスジェニック種子または穀物から製造した食品、飼料または飲料製品がある。この食品、飼料、または飲料は、小麦粉、練り生地、パン、パスタ、クッキー、ケーキ、シックナー、ビール、麦芽飲料、または食品添加物であってよい。この食品、飼料、またはビール製品は、低下したアレルゲン性および/または増大した消化性を持ち得る。さらに、練り生地製品は、同種の非トランスジェニック種子または穀物から製造した練り生地に比して、増大した強度および体積を持ち得る。この食品、飼料、または飲料は超消化性タンパク質および/または超消化性澱粉を持ち得る。この食品、飼料、または飲料は低アレルゲン性となり得る。上記の態様は当分野で知られるように、本発明に係る酵素を外部から添加することによっても達成できる。チオレドキシンによる小麦および乳中のジスルフィドタンパク質アレルゲンの還元がこれらのアレルゲン性を減少させることが示されている。さらにチオレドキシン処理は、他のジスルフィドタンパク質と同様、乳の主要アレルゲン (β -ラクトグロブリン) の消化性を増大させる。食品に外部からチオレドキシンを添加することの利点についてのより詳細な議論は米国特許No. 5,792,506に示されており、これを引用により特にここに含める。この組成物および方法は本発明に係るTR変異体を用いて増強することができる。

【0270】

本明細書で論ずるように、本発明に係るタンパク質を使用して食品および飼料中のタンパク質のアレルゲン性を低下させることができる。例えば、チオレドキシンとチオレドキシンレダクターゼならびに外部から加える処理としてのNADPHの使用に関するUS6190723およびその中の参考文献を参照されたい（これは引用により特にここに含める）。感作したイヌを用いて実施した皮膚試験および給餌実験は、摂取前にそれらの食物を処理するとアレルゲン性が取り除かれまたは減少することを示した。したがって、アレルゲン性食品または飼料タンパク質のアレルゲン性を減少させる組成物および方法が本明細書に提供される。食品もしくは飼料タンパク質または該タンパク質もしくはタンパク質群を含有する食品もしくは飼料を、該タンパク質のアレルゲン性の減少にとって有効な、或る量のチオレドキシン、チオレドキシンレダクターゼ、および補助因子、即ちNADPH、NADHまたはそれらの組み合わせと接触させる。その後、接触させたタンパク質を動物または人間に投与でき、ここで、該動物または人間の示したアレルギー性症状が対照に比して減少する。このアレルギー性食品/飼料タンパク質は好ましくは牛肉、牛乳、卵、大豆、米および小麦タンパク質である。チオレドキシンおよびTR変異体を含有し、さらに補助因子を含有する可食食品/飼料製品もまた具体化できる。酵素は外部から添加することができ、または1もしくはそれ以上が単独もしくは融合物としてトランスジェニックの形でまたは本来存在している。可食食品は、当該処理が原因で好ましくは低アレルゲン性である。この食物製品はペットフードまたはベビーフードまたは調合乳であってよい。この食物製品は牛肉、卵、大豆、小麦または乳タンパク質を含むことができる。これは可食食肉製品であってよい。米国特許5792506およびその参考文献を引用によりここに含める。

【0271】

同様に、US6114504では、シスチンを含有する動物および植物タンパク質を還元し練り生地およびベーカリー製品の性質を改善するための組成物および方法を提供しているが、これは、練り生地の成分をチオール酸化還元タンパク質と混合して練り生地を作り、この練

り生地を焼いてベーカリー製品を作る工程を含む。本発明方法は好ましくは還元されたチオレドキシンを小麦粉と共に使用するが、これはより強い練り生地とより大きなパンの体積を与える。この方法および組成物は本発明に係るタンパク質を使用して増強される。グルテニンまたはグリアジントタンパク質を還元する方法は、該グルテニンまたはグリアジントタンパク質を含有する液体または物質にチオレドキシンを添加し；チオレドキシンレダクターゼ変異体および補助因子、即ちNADPH、NADHもしくはそれらの組み合わせによってチオレドキシンを還元し、そしてこの還元されたチオレドキシンによってグルテニンまたはグリアジントタンパク質を還元することによる。組成物は、グルテニンまたはグリアジントタンパク質、添加されたもしくは内因性チオレドキシン、添加されたもしくは内因性（トランスジェニック植物由来の場合）チオレドキシンレダクターゼ変異体、および添加された補助因子、即ちNADPH、NADHもしくはそれらの組み合わせを含有する。この方法は、水不溶性または水溶性の種子由来タンパク質の含有を減らすのに有用である。チオレドキシンをこのタンパク質を含有する液体または物質に添加し；チオレドキシンレダクターゼ変異体およびその補助因子、即ちNADPH、NADHまたはそれらの組み合わせによってチオレドキシンを還元することができる。

【0272】

本発明はさらに、食品および飼料タンパク質の超消化性を増大させるのに有用である。チオレドキシン還元によって食物タンパク質の消化性を増大させる組成物および方法を提供している米国特許5952034を参照されたい。この方法は本発明に係る酵素を使用することにより増強される。食品の消化性を増大させる組成物および方法は、食品を、その食品の消化性を増大させるのに有効な、或る量のチオレドキシン、チオレドキシンレダクターゼ変異体、およびその補助因子、即ちNADPH、NADHまたはそれらの組み合わせで処理し；所望によりこの処理した食品を動物または人間に投与し、それにより、対照と比較した該動物または人間の示す症状によって測定される該食品の消化性を増大させることを含む。この食品は、好ましくは乳または小麦または卵を含む。上の態様では、チオレドキシンレダクターゼ変異体をチオレドキシンとのタンパク質融合物として提供できる。

【0273】

本発明に係る組成物はさらなる用途が見出せる。チオレドキシンおよびその他の酸化還元剤、例えばPDIはストレスと損傷に対する防御に有用であることが知られている。したがって、本発明に係る組成物は係る治療のための酸化還元剤組成物を増強するために使用できる。或る態様では、TR変異体を、ニトロソ化ストレスの操作に使用して、ニトロソ化ストレス防御をアップレギュレートする。US6359004を参照されたい。チオレドキシンはラジカル捕捉剤として働き、よってフリーラジカルに関連する疾病および状態は、好ましくはチオレドキシンと組み合わせた、TR変異体で治療され得る。したがって或る側面では、本発明は、眼疾患、例えば白内障の予防または治療のための組成物および方法を提供する。別の側面では本発明は、酸化的ストレスにより惹起されるまたは構成要素として酸化的ストレスを有する疾病の予防または治療に関するものである。例えば米国特許6379664を参照されたい。或る態様では、TR変異体を好ましくはチオレドキシンと組み合わせて含有する本発明に係る組成物の白内障阻害有効量に眼を接触させることにより、眼の白内障の形成を阻害または逆転させるための組成物および方法が提供される。別の態様では、TR変異体および補助因子と組み合わせたチオレドキシンの眼内注射が網膜の光酸化ストレスを抑制し、そして治療的戦略として網膜の光損傷を防止する。別の態様では、チオレドキシン活性を含む本発明組成物は、急性肺損傷で誘発される虚血-再灌流と酸化的ストレスを治療または最小化するのに有用である。その結果、特に末期肺疾患、例えば嚢胞性線維症、気腫、肺線維症、および肺高血圧症患者の肺移植に用途が見出せる。本発明に係る組成物は移植用臓器の完全性を維持するための保存用組成物としての用途が見出せる。別の態様では、TR変異体と組み合わせたチオレドキシンは一次培養ニューロンのインビトロ生存を促進する。さらに該組成物は、周縁部に神経防御効果を提供して、限局性脳虚血の際の神経損傷を緩和する。該組成物はさらに、神経損傷後またはこれに由来する運動ニューロンの防御および改善を提供する。別の態様では、本発明に係る組成物は虚血-再灌流損傷

から網膜を防御する。熱傷もまた本発明に係る組成物で治療できる。チオレドキシンおよびTR変異体は、迅速な抗酸化剤防御を提供し、凝固プロセス、細胞増殖、ならびに細胞防御と密接に結びついた細胞外過酸化物の状態の制御および熱傷の際の創傷治癒を改善する。最後に、本発明に係る組成物は、エプスタイン-バーウイルス（EBV）感染を制圧する良好な候補物質であるチオール-抗酸化物質を提供する。

【0274】

TR変異体は、チオレドキシンレダクターゼによりアスコルビン酸へと還元される有害なアスコルビンフリーラジカルおよびデヒドロアスコルビン酸塩を除去することによって直接的な利益が提供できる。したがってTRは直接的抗酸化効果および治療を提供する。該組成物は所望により補助因子を含有できる。

10

【0275】

本明細書に記載の疾病および状態において、TR変異体を単独でまたはチオレドキシンもしくは他の酸化還元剤および補助因子と組み合わせて供給できる。TR変異体は宿主の酸化還元剤と共に、または外部から添加した酸化還元剤と共に作用することができる。

【0276】

上記の発明を使用する方法をより詳細に描写するため、そして本発明の様々な態様の実施にとって考え得る最良の形態を開示するため、以下の実施例を供する。これらの実施例は本発明の真の範囲を限定するものでは決してなく、例示目的のために供するものである事が理解できるであろう。U.S.S.N. 60/289,029（2001年5月4日出願）、U.S.S.N. 60/370,609（2002年4月5日出願）、およびDesjarlaisおよびMuchhalによる「Novel Nucleic Acids and Proteins with Thioredoxin Reductase Activity」なる標題の仮出願（2002年4月29日出願、出願番号未定）を包含する本明細書に引用した全ての参考文献を、引用によりここに含める。

20

【実施例】

【0277】

実施例1

コンピューターによる変異体タンパク質の設計

概観

改良されたNADH-依存TR活性を持つ変異体を創出するための最初のPDA(登録商標)設計戦略を下に詳説する。一言で言えば、E coliおよびArabidopsis酵素由来の構造情報、ならびに補助因子のコンホメーション多様性を用いて、それぞれ～2000の組み合わせ成員を有する2つの異なるライブラリー（以後TR-1およびTR-2と称する）を設計した。

30

【0278】

スカホールドタンパク質として使用する野生型TR遺伝子：

1) pET29aにクローニングされたArbidopsis NTR1遺伝子。コードされているタンパク質はN末端Sタグを持つ。このタンパク質はBL21-S1細胞（塩誘導）またはBL21-Star（IPTG誘導）を用いて発現させ、BugBuster HTを用いて溶菌できる。

2) チオレドキシンj。pDEST-14で合成およびクローニングされ、BL21-S1-Starで発現されたコドン最適化遺伝子。Nおよび/またはC末端Hisタグを付けたものを作製した。C末端Hisタグを付けたTRxを動態決定に使用するためアフィニティークロマトグラフィーで精製した。

40

【0279】

検定：412nmにおけるDTNBの還元生成物形成の連続検出に基づく動態検定。

【0280】

「ヒット」の同定および動態特性決定に使用したスクリーニング戦略のさらに詳細な概観を図4に示す。

【0281】

全ての動態特性決定ならびに第二および第三階層のスクリーニングのため、精製タンパク質を使用した。必要量の精製タンパク質を産む高スループット法を個別に開発するか、または現存する商業的プロトコルから改変した。これらの方法の片鱗を図5に示す。高スル

50

ーブット培養、誘導、発現、タンパク質精製および酵素特性決定に使用した詳細なプロトコルを下に記載する。

【0282】

PDA（登録商標）設計された変異体についての基準点を規定するための、NADHおよびNADPH両基質に関する、精製WT NTR-1酵素（非修飾）の動態パラメータ（ K_m および K_{cat} ）。WT酵素は、天然（NADPH）補助因子についてNADHよりも～4倍高い K_{cat} を持つ（TRタンパク質1 μg を使用する V_{max} に相当）。さらにNADHの K_m はNADPHに比して～50倍高い。WT酵素についてのデータを図6に示す。

【0283】

位置指定突然変異誘発および反復PCRの標準的分子生物学的方法を用いてTRライブラリーを構築した。特異的に突然変異を受けた遺伝子セグメントを表す組み合わせ断片を特異的制限酵素を用いて合した。これらのライブラリーの品質をランダムに選んだクローンの配列および発現分析から評価した。TR-1とTR-2についてのこれらの詳細をそれぞれ図7および8に示す。これらの組み合わせライブラリーに加え、これら2つのライブラリー各々のための個別的C領域の組み合わせ（TR-1について24、TR-2について48）をWTバックボーンで合成し、PDA（登録商標）により同定したこの重要領域の効果を評価し、以降これらのクローンを、TR-3およびTR-4の個別的成員と共に「規定クローン」と称する（下記参照）。

10

【0284】

この2つのライブラリーのコンピューターに関連する記載を図9AおよびBに示す。設計された位置（橙色）および適当なコンホメーションを持つ結合させた補助因子（青または黄色）を明らかにしている。

20

【0285】

これら2つのライブラリーに加えて、さらなる戦略を探索するため幾つかの極めて小さなライブラリーを作製した。TR-3は18の成員を持ち、TR-2スクリーニング由来の最良クローンの結果に基づく細かな調整アプローチとして設計した。TR-4は16の成員を持ち、TRおよびAhpF配列の配列アラインメントに基づくものであった。AhpFは、TRと類似の活性であるNADH依存ペルオキシレドキシシレダクターゼをコードしている。

【0286】

これら4つのライブラリーのスクリーニングから得られた結果の要約を図10に示す。

【0287】

TR-1ライブラリーのスクリーニングは、WT NTR-1と比較して、NADHを補助因子とする著しく改善されたTR活性を持つクローンを何ら同定しなかった。これは「正しくない」補助因子のコンホメーションを用いた結果であろう。

30

【0288】

TR-2ライブラリーは著しく改善されたNADH依存活性を有する幾つかのクローンを持っていた。異なるC領域配列を持つ2つの最良変異体は「RYN」および「RFN」であった。設計された他の位置での突然変異はTR酵素の全体的性質に著しい影響を及ぼさなかった。以下のスライドはこれらの変異体の多くについての詳細な動態データを示すものである。

【0289】

異なる補助因子濃度におけるM-RYN、L-RYNおよびWTの動態パラメータおよび活性をそれぞれ図11AおよびBに示す。これらの変異体はいずれもWTと比較して著しく高いNADH依存活性を有する。さらにこれらは著しく低下したNADPH依存活性を有する。これを「補助因子スイッチ」と称する。2.5mMおよびそれ以上の補助因子濃度では、これらPDA（登録商標）設計されたNTRの両者は、NADHが補助因子であるWT NADPH活性の>50%を有する。

40

【0290】

これらのクローンの配列アラインメントおよび設計観点からの相対的コンピューター等級付けを図17Aに示す。

【0291】

RYNおよびRFNクローン中のNの存在が、可能性あるグルコシル化部位を作り出した。この部位はこれらのクローンの活性プロファイルに著しい影響を及ぼすことなくPDA（登録商

50

標)を用いて「設計された (designed out)」。このためのデータと戦略を下に記載する。

【0292】

重大なRRRからRYNへの変化のコンピューターによる表示を図18に示す。

【0293】

C領域におけるRYNおよびRFNの組み合わせに加えて、REN、RLN、RRNの組み合わせもまたNADH依存活性を著しく改善した。RRN変異体はまた、そのWTレベルのNADPH依存活性を維持した。このデータを図12にまとめる。さらに、RRT、RYT、RLR、KYN、MYN、QYN C領域変異体もまた改善されたNADH依存活性を示した。

【0294】

これらのライブラリーのスクリーニング結果は、補助因子の特異性プロファイルを決定するための、C領域に存在する3個のRRR残基の重要性を強く示している。これらの位置各々における20のアミノ酸の可能な全組み合わせの重要性を扱うため、高複雑性ランダムRRRライブラリーを設計およびスクリーニングして、NADHを伴うそれらの活性について最良の変異体を同定した。3個のR位置の各々においてNNK縮重を有するオリゴヌクレオチドを使用して、32768の成員の理論的組み合わせの可能性を有するこのライブラリーを構築した。

10

【0295】

このライブラリーの小部分のみをスクリーニングしたところ、最良クローンの配列および活性分析は、最初のR位置でのRからWへの突然変異が最も興味深い活性プロファイルを持つことを示した。これは、芳香族アミノ酸の存在を示唆する天然に存在する殆どのNAD(P)H依存性酵素配列のバイオインフォマティクス分析からも実証される。これは本発明者等を、PDA(登録商標)シミュレーションの間、最初のRが芳香族アミノ酸であらざるを得ないPDA(登録商標)ライブラリーの設計へと導いた。これにより、R1-WおよびWXXと呼ばれる、より小さなさらなる2つのPDA(登録商標)ライブラリーの設計が導かれた。これらの設計のためのコンピューター戦略を下に述べる。

20

【0296】

これら全ての新ライブラリーデザインからの最良のヒットを、2つの補助因子がそれぞれ0.6および1.2mMの時の相対活性について分析した(精製酵素を使用)。それらの K_m と K_{cat} も決定し、データをそれぞれ図13AとBに示す。これらのクローンは「高度に改善された」NADH依存TR活性を持つ。それらの改善されたNADH活性に加え、変異体の幾つかは改善されたNADPH依存活性をも有する。これは本質的に、両補助因子についてより良い触媒効率を有するTR変異体の創出を示す。これはさらに、全変異体について数倍高いNADH K_{cat} 値に反映している。NADHについての K_m は、WRTがこの補助因子について2倍低下した K_m を持っていることを除けば、改善された変異体の殆どについて不変であった。R1-WおよびWXXライブラリーのいずれかに由来するこのリストの成員を図13Cに示す。R1-Wライブラリー由来の2つの最良クローンのコンピューターモデルを、活性に関して構造上の見通しを得るために図14に示す。

30

【0297】

TRのためのPDA(登録商標)設計プロセスをこのように確認した：

40

- ・1.2mMのNADHでWT NADPH活性の50%に等しいまたはこれ以上の活性を有する、5またはそれ以上の変異体。
- ・少なくとも1個の変異体が、0.6mM NADHにおいてもこの活性の重要条件に適合する。
- ・これらの変異体の多数が、NADPH活性についても良好な触媒効率を有する。
- ・最良の変異体はWTに比してNADHについて13倍良好な K_{cat}/K_m および2倍低い K_m を持つ。

【0298】

チオレドキシシンレダクターゼR1-Wライブラリー

新たな組のPDA(登録商標)シミュレーションを実施して、Xencorの発見した三つ組み残基の最初の位置に芳香族アミノ酸(F、Y、またはW)を使用することが、NADHおよびNADPHで活性レベルを調節する際に極めて重要であることを評価した(RYN変異体中Rの位置に対

50

応する)。この新たなシミュレーションは、少数のNAD(P)H利用酵素がこの位置に芳香族を含むという知見、および、芳香族とNAD(P)H上のアデニン環の間の積層相互作用の可能性によって誘導された。

【0299】

インビトロスクリーニング用の1296の変異体を規定する 20^{10} (10^{13}) の配列のシミュレーションを下に示したライブラリーから得た。補助因子結合にとって重大な残基の構造分析により、10個の位置を選択した。このシミュレーション結果の分析は、10個のうち6個の位置におけるアミノ酸の相違をサンプリングすることにより、適度なサイズの高品質ライブラリーが得られることを明らかにした。

【0300】

1つの位置はW対Rの状況で、6個の位置に相違がある第4のPDA（登録商標）ライブラリーを以下のように規定する：

L I R R R V I (wt)

L I W R T V I

A L A S I V

F V C N

E C

K

L

M

Q

S

【0301】

このライブラリーの高スループットスクリーニングは以下の高活性WXXクローンを生成した。これらのクローンを、第4のPDA（登録商標）組み合わせライブラリーを示すPDA（登録商標）シミュレーションを行うことによりコンピューターで等級付けた。

【0302】

このライブラリー中の1296の可能な配列から、非常に活性なWXXクローンが以下のようにコンピューターで等級付けられる：

LIWRTVI 13/1296 (等級/ライブラリーサイズ)

LIWLSVI 51/1296

LIWMSVI 26/1296

LIWRSVI 46/1296

【0303】

これらの等級付けは相対活性を予想しようとするものではないという事に留意されたい。この計算は、構造的に共存可能な補助因子結合ポケット多様性の最も大きな集合を、最も少ない数の配列で定義すべく設計された。このライブラリー成員の全ては、第4ライブラリー中に含まれる6個の位置にある理論上可能な 20^6 の配列組み合わせの上位0.001%にあり、 10^4 以上の集束効果を証明している。それはさらに、元のシミュレーションに含まれる 20^{10} の配列組み合わせに比して少なくとも 10^9 の集束効果を構成する。

【0304】

これらの等級付けは純粋にNADHとのシミュレーションされた相互作用に基づくものであることにも留意されたい。これらはNADPHのためのまたはこれに対立する酵素の特異性を考慮していない。このプロジェクトの目的はNADPH/NADH特異性を包含しなかったため、2つの補助因子-タンパク質複合体の比較モデル作製は実施しなかった。

【0305】

さらなる変異体

R1-ライブラリーの成功、ならびにシミュレーションおよび研究室でのスクリーニングの両方で第二および第三のR位置にかなりの多様性が観察された事に基づき、Xencorは、全ての可能なWXX組み合わせをサンプリングするための小複雑性（400）ライブラリーを構築した。このライブラリーの高スループットスクリーニングにより、NADHを使用する高活性の変異体、およびNADPHを使用する種々の活性の変異体をさらに幾つか発見するに至った。

【0306】

3個のRRR位置にのみ多様性を含む、このライブラリー由来の最良クローン5個を下に列挙する。このライブラリーの設計は先のPDA（登録商標）シミュレーションおよび実験結果の全てに直接影響を受けたが、このライブラリーはPDA（登録商標）シミュレーション自体に基づくものではなかった。したがってこれらの変異体についてのコンピューターによる等級付けは無い。

WIS

WFQ

WVR

WMG

WVG

【0307】

RYNチオレドキシンレダクターゼ変異体のコンピューター等級付け

Xencorによって構築されスクリーニングされた第二PDA（登録商標）組み合わせライブラリーを表すPDA（登録商標）シミュレーションを実施することにより、個々の「RYN」クローンをコンピューターで等級付けた。20⁸ (2.5 x 10¹⁰) 配列のシミュレーションによって下のライブラリーを得たが、これはインビトロスクリーニング用の2304の変異体を規定する。補助因子結合のための重要残基の構造分析により、8個の位置を選択した。

【0308】

8個の位置に多様性を持つ第二PDA（登録商標）ライブラリーを下のように規定する：

LIGDRRRS

QMSNKYTD

L QEN

LI

【0309】

このライブラリー中の2304個の考え得る配列から、野生型および高度活性RYNクローンを以下のように等級付ける：

LIGDRRRS (wt) 329

LIGDRYNS 339

LLGDRYNS 698

LMGDRYNS 920

【0310】

これらの等級付けは相対活性を予想しようとするものではないという事に留意されたい。この計算は、構造的に共存可能な補助因子結合ポケット多様性の最も大きな集合を、最も少ない数の配列において定義すべく設計された。このライブラリー成員の全ては、第2ライブラリー中に含まれる8個の位置にある理論上可能な20⁸の配列組み合わせの上位0.00001%にあり、10⁷以上の集束効果を証明している。

【0311】

これらの等級付けは純粋にNADHとのシミュレーションされた相互作用に基づくものであることにも留意されたい。これらはNADPHのためのまたはこれに対立する酵素の特異性を考

10

20

30

40

50

慮していない。このプロジェクトの目的はNADPH/NADH特異性を包含しなかったため、2つの補助因子-タンパク質複合体の比較モデル作製は実施しなかった。

【0312】

新規なチオレドキシンレダクターゼ変異体

低複雑性ライブラリー。RYN変異体での最初の成功は、このRYN変異体中のアミノ酸を洗練することによりこの変異体のさらなる最適化をXencorに遂行させる動機となり、下に示す極めて小さな18成員ライブラリーを導いた。

R R R

M Y N

F D

10

【0313】

このライブラリーのスクリーニングは、RFNの組み合わせが、過去に発見されたRYN変異体と同様の活性を持つことを明らかにした。PDA（登録商標）シミュレーションによれば、このクローンはこのライブラリーの7番目に等級付けられる（RYNは3番目に等級付けられる）。

【0314】

非グリコシル化変異体。RYNおよび関連変異体（RYDAFNASKIMQQ）に可能性あるN結合グリコシル化部位（コンセンサスN-X-[T/S]）が偶然に導入されたため、PDA（登録商標）シミュレーションを実行して、RYN変異体においてAsn（N）の2位下流にあるセリン（S）の置換によってこの可能性ある部位を消去できる可能性を評価した。このシミュレーションは、SerからAlaへの置換を包含する幾つかのアミノ酸置換が好ましいことを示し、その後Xencorはこれを実験的に作製および特性決定した。この1位置シミュレーション（NAX）においては、Alaは6番目に、そしてThrおよびSerはそれぞれ1番目および2番目に等級付けられた。実験データは、Ala置換はRYN変異体の活性に検出可能な効果を持たないことを示している。

20

【0315】

RYN-A (339/2304,6/20) （等級/元のライブラリーサイズ、等級/NAXライブラリーサイズ）
RFN-A (7/18,6/20)

30

【0316】

コンピューターによる戦略

一次目標：NADH対NADPHを効率的に利用するようarabidopsisチオレドキシンレダクターゼの活性を変換。

【0317】

戦略の基本的概略

I. 開始モデルを作製

E coli構造（1TDF）を使用して、NADP補助因子の座標を補助因子座標を包含しないarabidopsis構造（1VDC）の座標枠に「融合」させる。

【0318】

40

II. 作業用補助因子コンホメーションを規定。

- NADPからPを除去することにより直接誘導。
- 種々のNAD利用酵素からのNAD座標の重ね合わせにより間接誘導

【0319】

III. 組み合わせライブラリーの可能性を作製するためのPDAシミュレーションを実行。

- ライブラリー位置を規定。
- シミュレーションを実行。
- ライブラリーを作製。

【0320】

戦略の詳細な概略

50

I. 開始モデルの作製

A. 1VDC構造ファイルをプロセッシングして、この構造のためのより妥当な番号付け系を創出した（元のバージョンは、その番号がE coli構造に合致するよう非定型的番号付けフォーマットを含んでいた）。

【0321】

B. NADP座標を1TDFから1VDCに融合するための構造アラインメント

以下の残基由来のC-アルファを用いてアラインメントを得た：117、119、151-156、174-181、および242-244。これは19の合致した原子について0.48ÅのRMSDを与える（最大偏差0.89Å）。

【0322】

C. 最終モデルに関して最小化を行わなかったことに留意されたい。

【0323】

II. 作業用補助因子コンホメーションを規定。

A. 1TDFファイル内にあるNADP補助因子から単に磷酸基を除去することによって最初の補助因子コンホメーションを規定した。本発明者等はこのコンホメーションをNAD_TDFと称する。

【0324】

B. これに代わるNADコンホメーション

Adam Thomasonは、NAD補助因子を含む構造についてPDBをスキャンするPerlスクリプトを開発した。次いでこのスクリプトは、抽出されたPDBファイル由来のNADをレファレンスNAD_TDF上に完全または部分的に重ね合わせる。このようにして多数のNADコンホメーションを収集し（図19を参照されたい）、PDAシミュレーションにおける使用準備が整った。

【0325】

NAD_TDFコンホマーまたはNAD_GRBコンホマー（1GRB ヒトグルタチオンレダクターゼより）のいずれかを用いてシミュレーションを実施したが、これはNAD_TDFに対して最も低い全原子r.m.s.dを持っていた。100以上のNADコンホマーの目視検査は、NAD_GRB中に見出されるリボースパッカーはNAD_TDF中のそれよりも著しく多く、このコンホマーが低いエネルギーを持つことを示唆している。NAD_TDF中にコンホマーが稀であるという事は、このコンホマーがNADP座標から誘導されたという事実に由来するのかも知れない。

【0326】

C. ヒドロキシ回転異性体の状態

ヒドロキシ基の水素の向きは側鎖-補助因子相互作用、特に水素結合相互作用に関して著しい影響を及ぼし得る。ライブラリー1について、ヒドロキシ回転異性体の静止対を利用した。何故なら、シミュレーション当たり1つのリガンドの状態だけがXencorのPDA（登録商標）の実行の中に含まれるためである。続いて、リガンド状態の組み合わせ集合がこのシミュレーションに包含されるようSPAパッケージを開発した。ヒドロキシ回転異性体の向きの組み合わせ集合を作製するため、「makeligands」（makeligands.f90より）なる名称の支援プログラムも開発した。

【0327】

III. 組み合わせライブラリーを作製するためのPDAシミュレーション

A. ライブラリー位置を規定

この戦略は、TRRタンパク質と、特にアデニンリボース上にジオール基（これは磷酸が除去される際そのまま残される）を有するNADHのアデニン部分との相互作用を増強することである（図20を参照されたい）。

【0328】

B. ライブラリー1の計算 PDA（登録商標）により実施

PDA（登録商標）シミュレーションパッケージを用いて最初の組み合わせライブラリーを作製した。このパッケージでは、リガンドが「テンプレート」の一部として取り込まれ、それはシミュレーション当たりのリガンド状態の数を1に制限する。故に、アデニンジオール上のヒドロキシ回転異性体はこの計算の集合について随意であった。さらに、NADに

10

20

30

40

50

ついて電荷は創出されなかった。計算の第一の集合は、189位における幾つかのアミノ酸の可能性を含んでいた。続く全ての計算について、この位置における実体はヒスチジンに限定された。

【0329】

C. ライブラリー1の定義

ライブラリー1の原理は、(i) ORBITによって予想された残基の質（ORBITモンテカルロシミュレーションにより作製した確率表に基づく）；(ii) 構造に対する直感；および(iii) 多様なアミノ酸性質をサンプリングすることの重要性、の組み合わせに基づいていた。このライブラリーの最も興味深い側面は、127位に残基EDTを、195位にQEを、217位にEQを、そして255位にEを生ずる、側鎖と補助因子の間の可能性ある様々な水素結合相互作用である。殆どのNADH利用酵素がカルボキシ側鎖とアデニンジオールとの相互作用を含んでいることから、195位におけるQとEの予想は有望である。

10

【0330】

TRRライブラリー1:

127 LEDTA
165 IML
166 G
167 G
189 H
190 RYM
191 RQ
195 RYQE
217 SEQ
255 IE

20

【0331】

D. ライブラリー2の計算 SPAにより実施

種々の補助因子コンホメーションおよびサンプリング戦略を用いた幾つかのシミュレーションをライブラリー2の開発のために実施した。

【0332】

(i) シミュレーションの第一の集合を、重原子座標のためのNAD_TDF補助因子コンホメーションを用いて実施した。このコンホメーションと、アデニンジオール上の36(6x6)のヒドロキシ回転異性体の組み合わせを使用して、バックボーン集団またはsub-回転異性体サンプリング戦略のいずれかによるシミュレーションを実施した。

30

【0333】

(ii) シミュレーションの第二の集合は、重原子座標のためのNAD_GRB補助因子コンホメーションを用いて実施した。このコンホメーションと、アデニンジオール上の36(6x6)のヒドロキシ回転異性体の組み合わせを使用して、バックボーン集団またはsub-回転異性体サンプリング戦略のいずれかによるシミュレーションを実施した。

【0334】

E. ライブラリー2の定義

ライブラリー2の原理は、(i) SPAによって予想された残基の質（出力自由エネルギーマトリックスおよび異なるシミュレーション由来のマトリックスの比較に基づく）；(ii) 構造に対する直感；(iii) 多様なアミノ酸性質をサンプリングすることの重要性；および(iv) ライブラリー1のスクリーニングからのフィードバック、の組み合わせに基づいていた。全ての位置においてこのライブラリーには野生型残基が含まれていた。前と同様、このライブラリーの最も興味深い側面は、側鎖と補助因子の間の可能性ある様々な水素結合相互作用である。しかしながら、これらの計算には別の補助因子コンホマーを使用したため、127位に残基Qを、167位にSを、195位にTNを（図3A、B）、217位にDを、そして255位にEを生ずる、新たな相互作用の集合がSPAによって予想された。アデニンジオールのA02^{*}酵素との水素結合に対する予想された能力と、小さな動きがファンデルワールス衝突を

40

50

軽減するであろうという推測に基づき、S167 (図3C) が、高い自由エネルギー値にも拘わらず選択された。169位にある負の電荷の中和が補助因子の結合親和性の改善を助けるであろうという可能性に基づき、169位にさらなる残基Nを加えた (NはE coli TRRに見出されるため、これは保存的突然変異であることに留意されたい)。

【0335】

ライブラリー2中の殆どの残基をNAD_GRBによるシミュレーションに基づいて選択した。しかしながら、NAD_TDF補助因子コンホメーションを用いるSPA計算においてI195の傾向が高い事に基づき、I195を加えた。

【0336】

TRRライブラリー2:

10

【表2】

126	123	118	R				1
127	124	119	L	Q			2
128	125	120	S				1
164	161	150	V				1
165	162	151	I	M	L		3
166	163	152	G				1
167	164	153	G	S			2
168	165	154	G				1
169	166	155	D	N			2
170	167	156	S				1
189	186	175	H				1
190	187	176	R	K	Q		3
191	188	177	R	Y	E	L	4
192	189	178	D				1
193	190	179	A				1
194	191	180	F				1
195	192	181	R	T	N	I	4
196	193	182	A				1
216	213	202	S				1
217	214	203	S	D			2
218	215	204	V				1
254	251	242	A				1
255	252	243	I				1
256	253	244	G				1

20

30

40

2304

【0337】

検定

発現

1. pET29でクローニングされたNTRコード領域をBL21 Star (Invitrogen) 細胞で発現させる。本明細書に記載の体積は精製タンパク質>50ugを取得するために典型的な体積であるが、必要に応じてスケールアップまたはダウンすることができる。
2. 1.5ml CG + カナマイシン (100ug/ml)を入れた96深ウェルプレートにコロニーを接種し、適当な対照を接種する。37°C、250rpmで培養を一夜増殖させる。
3. 翌日、96深ウェルプレートの各々から5ml CG + カナマイシン (100ug/ml) を入れた4x

50

- 24ウェルに一夜培養200 μ lを接種する。30°C、250rpmで3時間増殖させる。
4. 残りの一夜培養からグリセロール保存菌を作製し、-80°Cで凍結する。
5. 5ml培養を1M IPTGで誘導し最終濃度1mMとする。30°C、250rpmで一夜増殖させる。
6. 翌日、最大速度 (Avanti J-20、5300rpm) で10分間遠心沈降させる。上清を廃棄し、ペレットを-80°Cで凍結またはS.tag精製法へと進めることができる。

【0338】

96ウェルプレートのためのS.Tag精製

(96試料 (細胞ペレットより; Novagen、cat# 69232-3))

S.Tagトロンビン精製キットはビオチニル化トロンビンを用いる特異な戦略を利用するものであり、ストレプトアビジンアガロースによる消化後にこの酵素を単純且つ特異的に除去することができる。標準プロトコルは、Sタンパク質アガロースとのバッチ型結合、洗浄、ビオチニル化トロンビンによる処理、およびストレプトアビジンアガロースによる捕捉を必要とし、溶液中に精製タンパク質が残る。

【0339】

キット構成要素

【表3】

構成要素	体積	1キット/24試料当たりの体積
Sタンパク質アガロース (50mM Tris-HCl (pH7.5) 中50%スラリー、150mM NaCl、1mM EDTA、0.02%アジドナトリウム)	2ml	167 μ lスラリー/試料
10X結合/洗浄緩衝液 (200mM Tris-HCl (pH7.5)、1.5M NaCl、1% Triton X-100)	3x5ml	1X1ml/試料 100ml
10Xトロンビン開裂緩衝液 (200mM Tris-HCl (pH8.4)、1.5M NaCl、25mM CaCl ₂)	3ml	1X400 μ l/試料 30ml
ビオチニル化トロンビン	50U (1.5U/ μ l)	25U (16.6 μ l) 1U (0.66 μ l)/試料
ストレプトアビジンアガロース (磷酸緩衝液 (pH7.5) 中50%スラリー、0.02%アジドナトリウム)	2x0.4ml	1.6mlスラリー 60 μ lスラリー/試料

【0340】

さらなる材料:

Whatman Unifilter、96ウェル、800 μ l (Fisher、cat# PF7700-2804)

Bug Busterタンパク質抽出試薬 (VWR、cat# 80500-208)

【0341】

プロトコル (発現培養5ml)

- 凍結ペレット (5ml) を室温で~30分間融解させる。
- Bug Buster HT 500 μ lを加え、攪拌してペレットを再懸濁し、室温で20分間振盪する。
- 最大速度または3000xgで20分間遠心分離する。可溶性タンパク質を含有する上清 (細胞溶解液) を新たなプレートに移す。
- 細胞溶解液150 μ lを精製用を使用し、残りを後の使用のために保存する。

150 μ lについて

Tris-HClおよびNaCl濃度を20mM Trisおよび150mM NaCl (pH7.5) に調節する。

150 μ l Bug Buster $\times 100$

10 μ l 1M Tris-HCl (最終濃度20mM) 1ml

15 μ l 5M NaCl (最終濃度0.15M) 1.5ml

325 μ l H₂O 32.5ml

500 μ l 合計 アリコート 350 μ l ミックス

【0342】

5. フィルタープレートの底にアルミニウムテープで封をする。
6. 広口チップを用いてSタンパク質アガロースミックス167 μ lを加える。
7. 溶解液（調節済み）をフィルタープレートに加え、アルミニウムテープでプレートに封をする。
8. 楕円振盪機上、室温で30分-1時間結合させる（プレートを側方に位置させる。タンパク質が変性する傾向があるため激しく攪拌してはならない）
9. 底からアルミニウムテープをはずし、減圧に付す。
10. 500 μ lの1X結合/洗浄緩衝液で2回洗浄し、減圧に付す。
11. $\sim$ 1Xスラリー体積=200 μ lで1Xトロンビン開裂緩衝液により2回平衡化し、ごく低い減圧に付す。
12. フィルタープレートの底にアルミホイルで再度封をする。
13. 1Xトロンビン開裂緩衝液およびビオチニル化トロンビンのミックスを作製する。

【0343】

マスターミックス

【表4】

試薬	24試料につき1キット	
	各々	X100
1Xトロンビン開裂緩衝液	80 μ l	8ml
ビオチニル化トロンビン (1.5U/ μ l)	0.66 μ l	66 μ l
アリコート		80.7 μ l

14. 管を、設定=5、振幅=4のマイクロミキサー上、室温で1-2時間穏やかに振盪する。
15. ストレプトアビジンアガロースのスラリー60 μ lを加える。
16. 楕円振盪機上、室温で10分間インキュベートする。
17. フィルタープレートの底からホイルの封を取り除く。
18. 500xgで2分間遠心分離する。
19. より多くのタンパク質を溶離するため1X開裂緩衝液80 μ lを添加し、500xgで2分間遠心分離する。
20. 同体積の50%グリセロールを加え、十分に混合し、一時的には4℃で保存し、長期保存用には-80℃で凍結する。

【0344】

BCA検定

BCAタンパク質検定試薬キット (Pierce、cat# 23227)

1. 標準および作業用試薬の調製
 - a. 標準（作業範囲は0.125 - 2 μ g/ μ l）

【表5】

管	希釈剤の体積 (μ l)	BSAの体積	最終BCA濃度 (μ g/ μ l)
A	0	300 μ lの保存液	2.000
B	125	375 μ lの保存液	1.500
C	325	325 μ lの保存液	1.000
D	175	175 μ lのB	0.750
E	325	325 μ lのC	0.500
F	325	325 μ lのE	0.250
G	325	325 μ lのF	0.125
H	400	100 μ lのG	0.025
I	400	0 μ l	0.000=ブランク

検定のために：各標準 $5\mu\text{l}$ + ddH₂O $20\mu\text{l}$ = 計 $25\mu\text{l}$

【0345】

b. 作業用試薬

試薬A 50ml を試薬B 1ml と混合。

* この作業用試薬は密閉容器中、室温で保存すると数日間安定である。

【0346】

2. 96ウェルプレートにおける試料の調製。

精製タンパク質 $5\mu\text{l}$ (精製法の工程20より)

ddH₂O $20\mu\text{l}$

よく混合する

10

【0347】

3. 検定法

a. $25\mu\text{l}$ の標準および試料を入れた各ウェルに作業用試薬 $200\mu\text{l}$ を加える。

b. プレート振盪機上で30秒間プレートを完全に混合する。

c. アルミホイルテープでプレートを覆う。

d. 37°C で30分間インキュベートする。

e. プレートを室温に冷却する。

f. プレート読み取り機で 562nm の吸光度を測定する。

【0348】

4. 標準曲線のプロットにExcelを使用し、試料のタンパク質濃度を決定する。

20

【0349】

5. 検定のためにタンパク質濃度を正規化する。

a. 濃度を確認するため正規化したタンパク質のタンパク質ゲルをランする。

b. SYPRO Orangeで30分間-1時間染色する (および/またはCoomassie blueで一夜)

c. Alpha Innotech Corporation Imager上でゲルを視覚化する。

d. Kodak 1D 3.5 Networkソフトウェアを用いてデンストメトリーを実施する。

【0350】

チオレドキシンレダクターゼ検定

1. 検定/ウェル当たり $50\mu\text{l}$ の最終体積を用いる384微量定量プレートで検定を開始する
：4x96ウェルプレートまでを1枚の384プレートに入れ、移し換える時に個々のパターンを記録しておく。

30

2. 正規化したタンパク質試料 $5\mu\text{l}$ を384微量定量プレートのウェルに移す。 1.2mM (またはその他の適当な濃度) のNADPHまたはNADH、および精製チオレドキシン基質 $2\mu\text{M}$ を検定に使用する。

3. 検定ミックスを調製する：

【表6】

	1rxn	300rxn
ddH ₂ O	35.1 μl	10.53ml
1M トリス pH 8.0	5.0 μl	1.5ml
0.5M EDTA	1.0 μl	300 μl
20mM DTNB	0.5 μl	150 μl
25mM NADPH又はNADH	2.4 μl	720 μl
100 μM 精製チオレドキシン	1 μl	300 μl
全	45 μl	13.5 ml

40

* 試験しようとする上清に検定ミックスを添加する直前にNADHまたはNADPHおよびチオレドキシン基質を加える。

4. Titertek Multidrop 384を用いて検定ミックス $45\mu\text{l}$ を加える。

50

5. 直ちにプレートにSpectramaxプレート読み取り機にかけ、データの収集を開始する。
 6. 動態パラメータ (KcatおよびKm) の測定のために一般的に以下の基質濃度範囲を使用した：

NADPH : 0.00、0.01、0.02、0.04、0.08、0.15、0.3、0.6、1.2、2.5、5.0および10.0 mM
 NADH: 0.02、0.04、0.08、0.15、0.3、0.6、1.2、2.5、5.0、10.0および20.0 mM。

各濃度について直線範囲の初期反応速度を決定した。このデータをGraphPad Prismソフトウェアを用いて解析し、標準ミカエリス-メンテンの式に適合させた。

【0351】

検定用途のためのチオレドキシンh (N末端Hisタグ) の調製

培養の調製

1. 2リットルの発現培養に、BL21 Star (DE3) 発現細胞中のチオレドキシンコードンopt.e-coli/pET28bの一夜培養を接種する。これは>100mgの精製タンパク質を生成する。
2. 増殖期間の後、1M IPTGで1mM IPTGの最終濃度に細胞を誘導する。30℃、250rpmで一夜増殖させる。
3. 翌日、2Lの培養を20本の50ml Falcon管に遠心沈降させ、培養100mlからのペレットのみを残して上清を廃棄する。ペレットを-80℃で凍結した後、上清の調製およびHisタグ精製を継続する。

10

【0352】

上清の調製：

1. 20のペレットを各々1mlのBugbusterに再懸濁し、250rpm、室温で20分間振盪する。
2. 細胞を遠心沈降させ、上清を50ml Falcon管内に合する。2-メルカプトエタノールを加えた同体積の2Xローディング緩衝液を加える。精製を進める。

20

【0353】

Hisタグタンパク質の精製：

1. Clontech TALON Superflow樹脂懸濁液6mlを4本の50ml Falcon管に加える。
2. 1Xローディング緩衝液30mlで樹脂を2回洗浄する。
3. 室温で穏やかに20分間攪拌することによりタンパク質を樹脂に結合させる。
4. 1Xローディング緩衝液30ml中、室温で10分間樹脂を洗浄する。
5. 1Xローディング緩衝液3mlに樹脂を再懸濁する。
6. 4本の管全てから得た懸濁液を、1本のClontech 10ml重力流カラム中に合する。
7. 1Xローディング緩衝液15mlで樹脂を洗浄する。
8. 250mMイミダゾール溶離緩衝液20mlに樹脂を再懸濁する。タンパク質を50ml管に2回溶出する。
9. 濾過によるイミダゾール除去および試料濃縮を継続し、または後の使用のために-20℃で凍結する。

30

【0354】

精製チオレドキシンの濾過および濃縮

1. 精製タンパク質試料をMillipore Ultrafree-4 Biomax 5Kフィルター管でランする。
2. 濾過洗浄緩衝液で試料を3回洗浄する。
3. 濃縮したタンパク質試料を合する。BCA検定を行って濃度を決定し、次いで50%グリセロール、20mM Tris-HCl pH8.0で100uMに希釈する。

40

【0355】

2Xローディング緩衝液

100mM NaPO4 pH 8.0

10mM Tris、pH 8.0

600mM NaCl

20mMイミダゾール

10% エチレングリコール

2mM 2-メルカプトエタノールを加えた2Xローディング緩衝液のために0.156ul/mlを添加する。

50

【0356】

250mMイミダゾール溶離緩衝液

50mM NaP04 pH 8.0

5mM Tris、pH 8.0

200mM NaCl

250mMイミダゾール

10% エチレングリコール

【0357】

濾過緩衝液 (イミダゾール除去のため)

50mM NaP04

10mM Tris、pH 8.0

200mM NaCl

10% エチレングリコール

ddH2O

【0358】

実施例2

変異体TRタンパク質による植物の形質転換

概観

オレオシン-TR融合タンパク質、オレオシン-TRレダクターゼ融合タンパク質またはオレオシン-ハイブリッドTRレダクターゼ/TRレダクターゼ融合タンパク質をコードしている遺伝子を、常套的組換えDNA技術を用いて植物細胞中に組み込むことができる。一般にこれは、オレオシン-TRレダクターゼ融合タンパク質、オレオシン-TRレダクターゼ融合タンパク質またはオレオシン-ハイブリッドTR/TRレダクターゼ融合タンパク質をコードしているDNA分子を上記のように発現系に挿入することを含む。

【0359】

育種

生産および品質にとって重要な他の性質と共にオレオシン-TR融合タンパク質、オレオシン-TRレダクターゼ融合タンパク質またはオレオシン-ハイブリッドTR/TRレダクターゼ融合タンパク質を発現する植物は、育種アプローチおよび当分野で既知の技術によって植物系統に組み入れることができる。オレオシン-TR融合タンパク質、オレオシン-TRレダクターゼ融合タンパク質またはオレオシン-ハイブリッドTR/TRレダクターゼ融合タンパク質を発現する植物が得られたならば、該アリルを遺伝学的に作り換えたりこれを植物中に導入する必要なく、伝統的な育種技術を用いてこのトランスジーンを商業的な変種に移す。

【0360】

母系組織が子孫を生ずる、アポミクシスの生殖能を有する植物は、オレオシン-R融合タンパク質、オレオシン-TRレダクターゼ融合タンパク質またはオレオシン-ハイブリッドTR/TRレダクターゼ融合タンパク質を発現するよう形質転換し、導入したアリルをアポミクシスの育種によって所望のバックグラウンドに維持することができる。

【0361】

TRおよびTRレダクターゼ遺伝子の単離ならびにインビトロ検定

或る態様では、Arabidopsis、小麦、仔牛のような哺乳動物起源およびE.coli由来のTR遺伝子を単離し、細菌発現ベクターを用いてE.coliで発現させ、得られたタンパク質生成物を精製することができる。別の態様では、ArabidopsisおよびE.coli由来のTRレダクターゼ遺伝子を単離し、E.coliで発現させ、そして精製することができる。さらに、TR/TRレダクターゼ遺伝子をMycobacterium lepraeから単離/取得し、E.coliで発現させ、精製できる。好ましい態様では、M.lepraeコドン、任意の与えられた宿主、例えばE.coli宿主細胞または植物種における最適化のために改変できる。多くの生物のためのコドン使用法の表が知られており利用可能であって、特定宿主のために作り変えられたコード配列のコドン最適化を可能にしている。

【0362】

別の態様では、改変された補助因子特異性を持つTRレダクターゼを、標的突然変異誘発またはランダム突然変異誘発を用いて製造し、補助因子結合部位における特異的突然変異について試験する (Shiraishi, et al. (1998) Arch Biochem Biophys 358 (1): 104-115; Galkin et al. (1997) Protein Eng 10(6): 687-690); Carugo et al. (1997) Proteins 28(1):10-28; Hurley et al. (1996) Biochemistry 35(18):5670-8; および/または有機溶媒の添加による (Holmberg et al. (1999) Protein Eng 12 (10): 851-856)。突然変異の決定は例えばMayoおよびDahiyatが開発したもの (Chem & Eng News October 6, 1997, pages 9-10) のようなコンピュータプログラムによって支援することができる。前記の参考文献は引用によりその全体をここに含める。

【0363】

異なるTRおよびTRレダクターゼの組み合わせをマトリックスで使用して、どのTRおよびTRレダクターゼの組み合わせがインビトロでの小麦貯蔵タンパク質および乳貯蔵タンパク質β-ラクトグロブリンの還元にも最も有効であるかを決定する。好ましくは、TRとTRレダクターゼの組み合わせを試験する。これらの実験はDel Val et al. ((1999) Jnl Allerg Clin Immunol 103:690-697) の記載に従って実施する。同系交配の高IgE応答アトピー犬を取得し、乳および小麦を含有する食物調製物の市販抽出物で感作することによってさらに準備する。I型過敏性反応を用いて皮膚試験を実施する。皮膚試験の直前にエバンスブルー色素を静脈内注射する。小麦粥、全牛乳抽出物および純粋なβ-ラクトグロブリンのアリコートを経内注射する。各々の青色スポットの2つの垂直直径を評点付けることにより、皮膚試験を盲検により読み取る。オレオシン-TR、オレオシン-TRレダクターゼおよびこれらの組み合わせがアレルギー反応に影響を及ぼす能力を、外因性NADPHまたはNADHの存在下または不在下で測定する。

【0364】

植物発現ベクターの構築

Arabidopsis TRおよびTRレダクターゼ遺伝子の配列が公表されており (Rivera-Madrid et al. (1995) Proc Natl Acad Sci USA 92:5620-5624; Jacquot et al. (1994) J Mol Biol 235:1357-1363)、これらの遺伝子をPCRで単離できる。

或る態様では、Arabidopsis TRおよびTRレダクターゼ遺伝子の両者を、オレオシンのNおよびC末端の両方に翻訳により融合させる。このオープンリーディングフレームは、植物中での発現のための適当なプロモーターおよびターミネーター配列の転写調節下にある。好ましい態様では、ファセオリンプロモーターおよびターミネーター配列を用いてArabidopsis TR (ATR) およびArabidopsis TRレダクターゼ (ATRR) 構築物を創出する。

【0365】

Arabidopsis中での発現

或る態様では、オレオシン-ATRおよびオレオシン-ATRR発現構築物の初回試験のためのモデル系としてArabidopsisを使用する。Arabidopsisの種子は、収穫種、特にTRの商業生産に利用できる油料種子収穫種に極めてよく似たオレオシン被覆油体を含有している。Arabidopsis中でのオレオシン-TRおよびオレオシンTRレダクターゼの発現を利用して、油体中のオレオシンTRおよびオレオシンTRレダクターゼ融合物を得、これらの融合タンパク質が生物活性であるかどうかを決定する。オレオシンに対するTRおよびTRレダクターゼ両方のNおよびC両末端融合物を作製し試験する。さらなる態様では、M.leprae由来の天然TR/TRレダクターゼ融合遺伝子に対するオレオシン融合物を試験する。これらの融合タンパク質の蓄積を、オレオシンおよび/またはTRおよびTRレダクターゼに特異的な抗体を利用するウェスタンブロッティングを用いて定量する。形質転換されたArabidopsis植物が再生および生育するのに要する時間、およびトランスジェンの発現と種子中に発現された生成物の蓄積を測定するのに要する時間が、殆どの収穫種よりはるかに短いため、Arabidopsisがこの目的にとって有用である。

【0366】

植物発現ベクターの構築

植物発現ベクターは、小麦由来のTR遺伝子、仔牛のような哺乳動物起源のTR遺伝子、E.co

10

20

30

40

50

li由来のTR遺伝子；E.coli由来のTRレダクターゼ遺伝子；M.leprae由来のTR/TRレダクターゼ遺伝子を包含する（但しこれらに限定される訳ではない）、TRおよびTRレダクターゼをコードしているその他の遺伝子を用いて構築する。これらの遺伝子のいずれかまたは両方をオレオシンのNおよびC両末端と翻訳により融合させる。このようないかなる構築物のオープンリーディングフレームも、適当なプロモーターおよびターミネーター配列の転写調節下にある。好ましい態様では、ファセオリンプロモーターおよびターミネーター配列を用いてTR'およびTRレダクターゼを目的とする植物発現ベクターを構築する。さらに好ましくは、ファセオリンプロモーターおよびターミネーター配列を用いてTR'およびTRレダクターゼ'を目的とする植物発現ベクターを構築する。

【0367】

ベニバナにおける発現

当業者に既知の方法を使用してベニバナを形質転換するために上記の植物形質転換ベクターを使用する。好ましい態様では、BakerおよびDyer (Plant Cell Rep (1996) 16:106-110) の開示した方法から適合させた方法によってベニバナを形質転換する。発現はノーザンおよびウェスタンブロッティングを用いて検定する。小麦貯蔵タンパク質および乳貯蔵タンパク質β-ラクトグロブリンを還元するTR'およびTRレダクターゼ'構築物の能力を試験する。各構築物当たり最低25の個別に形質転換させたトランスジェニックベニバナ植物を作製する。全てのトランスジェニック標的収獲植物をオレオシン-TR'およびオレオシン-TRレダクターゼ'の発現について試験する。この分析結果は、どの形質転換事象が最も高いそして/または最適なTR'またはTRレダクターゼ'活性をもたらすかを示す。この構築物で形質転換したトランスジェニック系統をさらなる分析に付す。TR'およびTRレダクターゼ'の量を定量的ウェスタンブロッティング分析を用いて決定する。オレオシン融合物の比活性を、E.coliで生産された「遊離の」TR'およびTRレダクターゼ'の比活性と比較する。

【0368】

最も高い発現を行った植物系統を繁殖させる。安定な遺伝子型を有する同型接合体および二倍体植物を作製して、後続世代における安定なトランスジーン遺伝形質を確実とする。

【0369】

ビオチニル化TRの製造

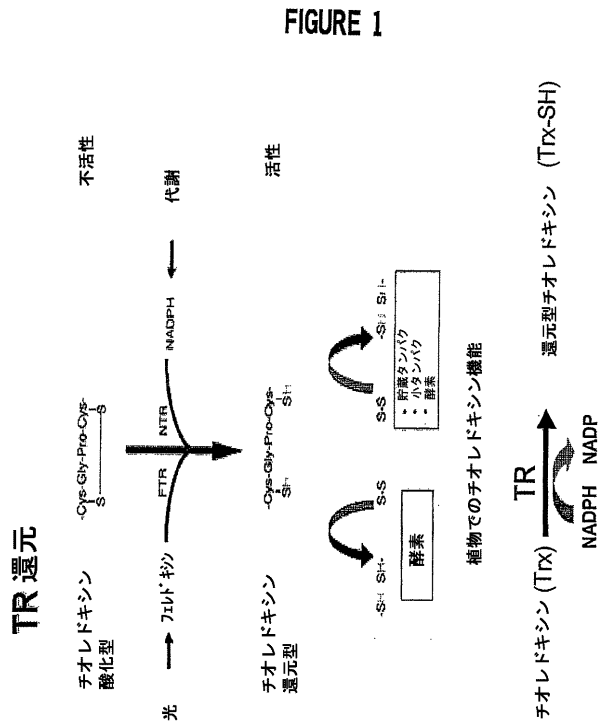
或る態様では、ビオチニル-N-ヒドロキシスクシンイミドエステルのような化学物質を用いてリジン残基を化学修飾することにより、TRをインビトロでビオチニル化することができる。これに代わる態様では、E.coliアセチル-CoAカルボキシラーゼのビオチンカルボキシ担体タンパク質由来のビオチンドメインペプチドを利用して、インビボの位置特異的ビオチニル化を、Smith et al. ((1998) Nuc Acid Res 26:1414-1420) の記載に従って利用できる。E.coli宿主内因性ビオチニル化機構 (BIOTRX) によりインビボビオチニル化が可能な組換えチオレドキシンを、E.coli trxAの5'末端に23アミノ酸ビオチニル化認識ペプチドをコードしているオリゴヌクレオチドをフレーム内挿入することによって構築し、構築物pBIOTRXを創出する。pBIOTRXプラスミドを含む細胞を外因性ビオチンの不在下に増殖させ、BIOTRXタンパク質の量および溶解度を決定する。総細胞タンパク質の10%までがBIOTRXタンパク質であることが判明し、一方少量のトリチル化ビオチンがBIOTRXタンパク質中に取り込まれ、固定化アビジンまたは固定化アビジン-アルカリホスファターゼに対するBIOTRXの結合は低い。pBIOTRXで形質転換された細胞の誘導前培地へのビオチン10 μg/mlの添加は、全体的なビオチン取り込みの程度の改善をもたらす。

【0370】

ビオチニル化油体-TR混合物の製造

アビジンまたはストレプトアビジンを用いてビオチニル化TRをビオチニル化油体に結合させる。精製したビオチニル化TRを異なる比率でビオチニル化油体と混合する。小麦のアレルゲン性を低下させ練り生地の品質を改善するこれら混合物の有効性を、乳製品のアレルゲン性を低下させるこれら混合物の有効性と共に試験する。対照は、野生型ベニバナ油体およびTRと混合した（但し結合していない）野生型ベニバナ油体を包含する。

【図 1】



【図 2 A】

チオレドキシンの還元 (Tx-SH)

FIG 2A

A1	A2	S2	A3	A4	A5	S3	A6	種
Q39243	S170	A171	17X	H189	R190	R191	3X	R195 アラビドプシス
Q39242	S220	A221	17X	H239	R240	R241	3X	R245 アラビドプシス
Q22229	T240	A241	17X	V259	R260	R261	3X	R265 アラビドプシス
P09625	T156	A157	17X	H175	R176	R177	3X	R181 E. coli
P29509	S164	A165	17X	V183	R184	K185	3X	R189 コウボ
P38816	S188	A189	17X	V207	R208	K209	3X	R213 コウボ
Q17745	V220	S221	17X	V239	R240	S241	3X	R245 C. elegans
Q9N2K1	V220	S221	17X	V239	R240	S241	3X	R245 C. elegans
Q9N2H3	V362	S363	17X	V381	R382	S383	3X	R387 C. elegans
Q9VNT5	V223	G224	17X	V242	R243	S244	3X	R248 ショウジョウバエ
O62768	V201	A202	17X	V220	R221	S222	3X	R226 ウシ
Q9N2I8	V216	A217	17X	I235	R236	S237	3X	R241 ウシ
Q16881	V201	A202	17X	V220	R221	S222	3X	R226 ヒト
Q95840	V229	A230	17X	M248	R249	S250	3X	R254 ヒト
Q9U8S9	V201	A202	17X	V220	R221	S222	3X	R226 ヒト
Q9U8T9	V201	A202	17X	V220	R221	S222	3X	R226 ヒト
Q9U8U8	V226	A227	17X	M245	R246	S247	3X	R251 ヒト
Q9N8W6	V281	A282	17X	V300	R301	S302	3X	R306 ヒト
Q9N8W7	V229	A230	17X	M248	R249	S250	3X	R254 ヒト
Q9P101	V279	A280	17X	V298	R299	S300	3X	R304 ヒト
Q9P2Y0	V159	A200	17X	M218	R219	S220	3X	R224 ヒト
Q9H2Z5	V228	A229	17X	M247	R248	S249	3X	R253 ヒト (ミトコンドリア)
Q9H475	V253	A254	17X	V272	R273	S274	3X	R278 ヒト
Q9H499	V315	A316	17X	V334	R335	S336	3X	R340 マウス
Q9C8V5	V201	A202	17X	V220	R221	S222	3X	R226 マウス
Q9C8Z5	V317	G318	17X	V336	R337	S338	3X	R342 マウス
Q9JH7	V229	A230	17X	M248	R249	S250	3X	R254 マウス
Q9JH74	V233	A234	17X	M252	R253	S254	3X	R258 マウス
Q9JH85	V225	A226	17X	M244	R245	S246	3X	R250 マウス
Q9JH86	V201	A202	17X	V220	R221	S222	3X	R226 マウス
Q9JH99	V201	A202	17X	V220	R221	S222	3X	R226 ラット
Q9JH34	V201	A202	17X	V220	R221	S222	3X	R226 ラット
Q9R1I3	V201	A202	17X	V220	R221	S222	3X	R226 ラット
Q9Z0J5	V231	A232	17X	M250	R251	S252	3X	R256 ラット
Q9MYT8	V201	A202	17X	V220	R221	S222	3X	R226 プタ

【図 2 B】

FIG 2B

グルタチオンレダクターゼ

A1	A2	S2	A3	A4	A5	S3	A6	種
P42770	I271	A272	17X	I290	R291	Q292	4X	R297 アラビドプシス
P48641	I214	A215	17X	F233	R234	K235	4X	R240 アラビドプシス
P48642	I211	A212	17X	V230	R231	K232	4X	R237 イネ
O64409	I83	A84	17X	I102	R103	Q104	4X	R109 トウモロコシ
P06715	I178	A179	17X	V197	R198	K199	4X	R204 E. coli
O01412	I182	A183	17X	I201	R202	K203	4X	R208 O. volvulus
P41921	I208	G209	17X	I227	R228	G229	4X	R234 コウボ
P91938	I192	G193	16X	V210	R211	S212	3X	R216 ショウジョウバエ
P00390	I242	A243	17X	I261	R262	H263	4X	R268 ヒト
P47791	I220	A221	17X	I239	R240	H241	4X	R246 マウス
P70619	I236	A137	17X	I155	R156	H157	4X	R162 ラット

【図 2 E】

FIG 2E

コファクター-特異性

A1	A2	S2	A3	A4	A5	S3	A6	種	特異性
TR	S	A	17X	H	R	R	3X	R	アラビドプシス
TR	T	A	17X	H	R	R	3X	R	E. coli
GR (wt)	I	A	23A13X	V	R	K	H1X	R	E. coli
GR (mut)	I	G	3XG13X	E	N	F	D1X	P	E. coli
Op34	S	A	17X	E	Q	F	3X	Q	C. pasteurianus
AhpF	S	G	17X	E	F	A	3X	K	S. cyphitricum

【図 2 C】

FIG 2C

チオレドキシンの還元 (Tx-SH)

A1	A2	S2	A3	A4	A5	S3	A6
アラビドプシス	S (T)	A	17X	H (V)	R	R	3X
E. coli	T	A	17X	H	R	R	3X
コウボ	S	A	17X	V	R	K	3X
C. elegans	V	S	17X	V	R	S	3X
ショウジョウバエ	V	G	17X	V	R	S	3X
動物	V	A (G)	17X	V/M (I)	R	S	3X

【図 2 D】

FIG 2D

グルタチオンレダクターゼ

A1	A2	S2	A3	A4	A5	S3	A6
植物	I	A	17X	I (F, V)	R	K/Q	4X
細菌	I	A	17X	V (I)	R	K	4X
コウボ	I	G	17X	I	R	G	4X
ショウジョウバエ	I	G	16X	V	R	S	3X
動物	I	A	17X	I	R	H	4X

【図 3 - 1】

FIG 3A

配列番号 1

E. coli チオレドキシンレダクターゼの N 末端配列 (S₁)

```
GTTHSKLLILGSGPAGYTAAYYAARANLQPVLITGMEKGGQLTTTTEVENWPGDPNDLT  
GPLLMERMHEHATKPETRIIFDHINKVDLQNRPFRLNGDNGEYTTCDALIIATGASARYLG  
LPSEBAFKGRGVSACATCDGFFYRNQKVAVIGGGN
```

FIG 3B

配列番号 2

Bacillus subtilis チオレドキシンレダクターゼの N 末端配列 (S₁)

```
SEEKIYDVIIIGAGPAGMTAAVYTSRANLSTLMIERGIPGGOMANTEBDVENYPGFESILG  
PELSNMFEBHAKKFGAEYAYGDIKEVIDGKEYKVVKAGSKEYKARAVIIAAGAEYKKIGV  
PGEKRLGGRGVSYCAVCDGAFPGKGLVVVGGGD
```

FIG 3C

配列番号 3

Mycobacterium leprae チオレドキシンレダクターゼの N 末端配列 (S₁)

```
MNITPSAHETIHEVIVIGSGPAGYTAALYAARAQLTPLVFEGTSTFGGALMTTTEVENYPG  
FRNGITGPELMDDMREQALRFGAELRTEDVESVSLRGFIKSVVTAEGQTYQARAVILAMG  
TSVRYLQIPGEQELLRGRVSAACATCDGSFFRGQDLAVIGGGD
```

FIG 3D

配列番号 4

Sarccharomyces チオレドキシンレダクターゼの N 末端配列 (S₁)

```
VHNKVTIIIGSGPAHTAAIYLARAEIKPILYEGMMANGIAAGGQLTTTTEIENFPGFPGD  
LTGSELMDRMREQSTKFGTEIITETYSKVDLSKPKFLWTEFNEDAEPVTTDAIILATGA  
SAKRMLHPGEETYWQKGISACAVCDGAVPIFRNKPLAVIGGGD
```

FIG 3E

配列番号 5

Neurospora crassa チオレドキシンレダクターゼの N 末端配列 (S₁)

```
MHSKVVIIGSGPAHTAAIYLARAEIKPILYEGFMANGIAAGGQLTTTTEIENFPGFPGD  
IMGQELMDKMKQSERFGTQIISSETVAKVDSLARPFKYATEWSPEEYHTADSIIILATGAS  
ARRLHLPGEEKYWQNGISACAVCDGAVPIFRNKHLVVIGGGD
```

【図 3 - 3】

FIG 3L

配列番号 12

Neurospora crassa チオレドキシンレダクターゼの 第一内在配列 (S₂)

```
AEBA MYLT KYGSHVTVL
```

FIG 3M

配列番号 13

Arabidopsis チオレドキシンレダクターゼの 第一内在配列 (S₂)

```
MEEANFLTKYGSKVYII
```

FIG 3N

配列番号 14

Human チオレドキシンレダクターゼの 第一内在配列 (S₂)

```
MVRSILLRGFDQDMANKIGEHEHMEHGIIKFI
```

FIG 3O

配列番号 15

E. coli チオレドキシンレダクターゼの 第二内在配列 (S₃)

```
DGF
```

FIG 3P

配列番号 16

Bacillus subtilis チオレドキシンレダクターゼの 第二内在配列 (S₃)

```
DKL
```

FIG 3Q

配列番号 17

Mycobacterium leprae チオレドキシンレダクターゼの 第二内在配列 (S₃)

```
DEF
```

【図 3 - 2】

FIG 3F

配列番号 6

Arabidopsis チオレドキシンレダクターゼの N 末端配列 (S₁)

```
MNGLETHNTRLICIVGSGPAHTAAIYAARAELKPLLFEGWMANDIAPGGQLTTTIDVENF  
PGFPPEGILGVELTDKFRKQSERFGTTITFETVTKVDPSSKPFKLTDSKAILADAVILAT  
GAVAKRLSPVGSGBEASGGFWNRGISACAVCDGAAPIFRNKPLAVIGGGD
```

FIG 3G

配列番号 7

Human チオレドキシンレダクターゼの N 末端配列 (S₁)

```
MNGPEDLPKSYDYDLIIIGGSGGLAAAKEAAQYGGKVMVLDVFTPTPLGTRWGLGGTGV  
NVGCIPIFKLMHQAALLGQALQDSRNYGKWVEETVKHDWDRMIEAVQNHIGSLNWGYRVAL  
REKKVVYENAYGQFIGPHRIKATNNKGKEKIYSAESFLIATGERPRYLIGIPGDKEYCISS  
DDLFSLFYCPGKTLVVGASYVALBCAGFLAGIGLV
```

FIG 3H

配列番号 8

E. coli チオレドキシンレダクターゼの 第一内在配列 (S₂)

```
VEEALYLSNIASEVHLI
```

FIG 3I

配列番号 9

Bacillus subtilis チオレドキシンレダクターゼの 第一内在配列 (S₂)

```
VEEGVYLTRFASKVTIV
```

FIG 3J

配列番号 10

Mycobacterium leprae チオレドキシンレダクターゼの 第一内在配列 (S₂)

```
MEEALFLTRFARSVTLV
```

FIG 3K

配列番号 11

Sarccharomyces チオレドキシンレダクターゼの 第一内在配列 (S₂)

```
CERAEFLTKYGSKVFMIL
```

【図 3 - 4】

FIG 3R

配列番号 18

Sarccharomyces チオレドキシンレダクターゼの 第二内在配列 (S₃)

```
DHL
```

FIG 3S

配列番号 19

Neurospora crassa チオレドキシンレダクターゼの 第二内在配列 (S₃)

```
DKL
```

FIG 3T

配列番号 20

Arabidopsis チオレドキシンレダクターゼの 第二内在配列 (S₃)

```
DAF
```

FIG 3U

配列番号 21

Human チオレドキシンレダクターゼの 第二内在配列 (S₃)

```
VPI
```

FIG 3V

配列番号 22

E. coli チオレドキシンレダクターゼの C 末端配列 (S₄)

```
AEKILIKRLMDKVENGNII LHTNRTLSEVTGDMGVGTGVRLRDTQNSDNIESLDVAGLFPV  
AIGHSPNTAIFEGQLELENGYIKVQSGIHGNATQTSIPGVFAAGDVMDDHIYRQAITSGAT  
GCMAALDAERYLDGLADAK
```

FIG 3W

配列番号 23

Bacillus subtilis チオレドキシンレダクターゼの C 末端配列 (S₄)

```
AQSILQARAFDNEKVDPLWNKTVKEIHEENGKVGNVTLVDVTGTGESESPKTDGVFIYIGM  
LPLSKPFENLGIITNEEGYIETNDRMETKVBGIFPAAGDIREKSLRQIVTATGDSIAAQSV  
QHVVBEELQETLTKLTK
```


TR ライブラリー 1



FIGURE 9A

TR ライブラリー 2



FIGURE 9B

【図 9 A】

1	L	E	D	T	A
2	I	M	L		
3	R	Y	M		
4	R	Q			
5	R	Y	Q	E	
6	S	E	Q		
7	I	E			

2160

【図 9 B】

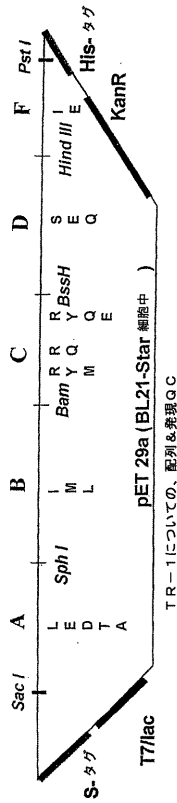
1	L	Q			
2	I	M	L		
3	R	K	Q		
4	R	Y	E	L	
5	R	T	N	I	
6	S	D			
7	I				
8	G	S			
9	D	N			

2304

NTR-1 ライブラリー #1 (iR-1)

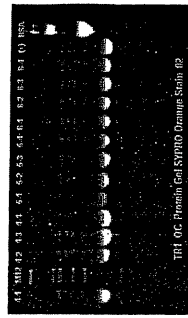
理論的多様性 = 2160

=それぞれ90員と定義C領域での24ミニライブラリーOCによるA、B、D&F、帰納的PCRによるC領域、ライゲーションによる結合(3片)



L-21 (26)	L-31 (38)	S-23 (90)	L-47 (60)
E-20 (24)	M-36 (44)	E-24 (90)	E-32 (40)
D-8 (10)	L-14 (18)	Q-32 (40)	
T-20 (24)			
A-18 (17)			
84	81	79	79

配列多様性は、よく現れる



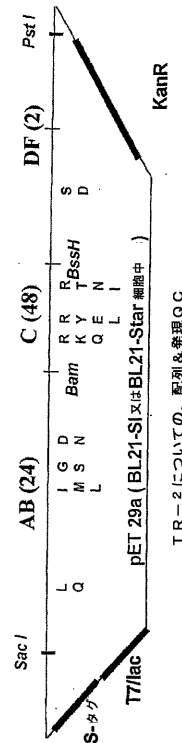
8400 クロローンに基づく
発現... ~ 90%発現 (7/84NS)

FIGURE 7

NTR-1 ライブラリー #2 (TR-2)

理論的多様性 = 2304

=それぞれ48員と定義C領域での48ミニライブラリーOCによるA、B&D、帰納的PCRによるC領域、ライゲーションによる結合(3片)



L 38	I 27	G 40	D 44	S 37
Q 46	M 33	S 45	N 41	D 49
		L 25		



Star vs SI 発現

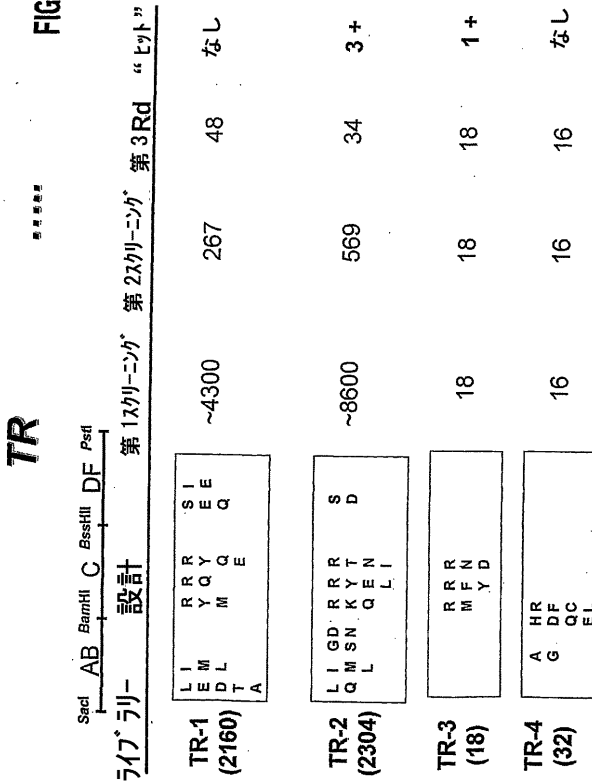
>90%クロローンは、
可溶性TRタンパク質を発現する

OW

FIGURE 8

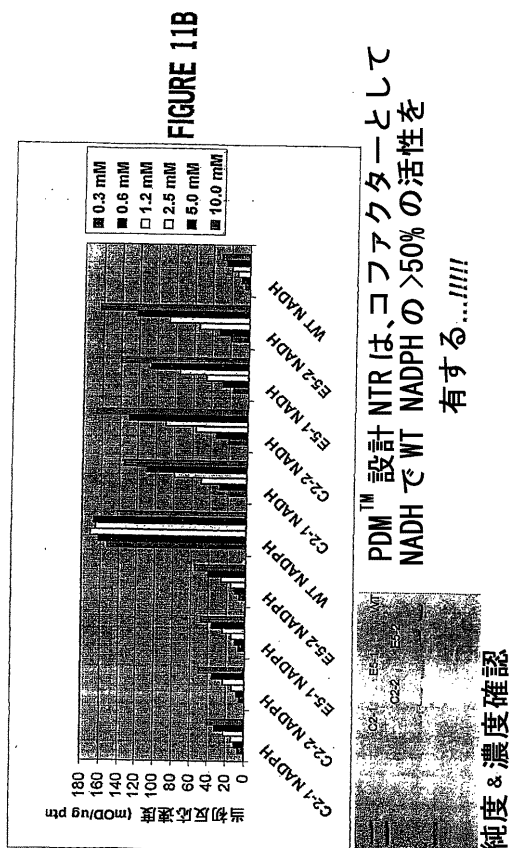
【图 10】

FIGURE 10



【 図 1 1 B 】

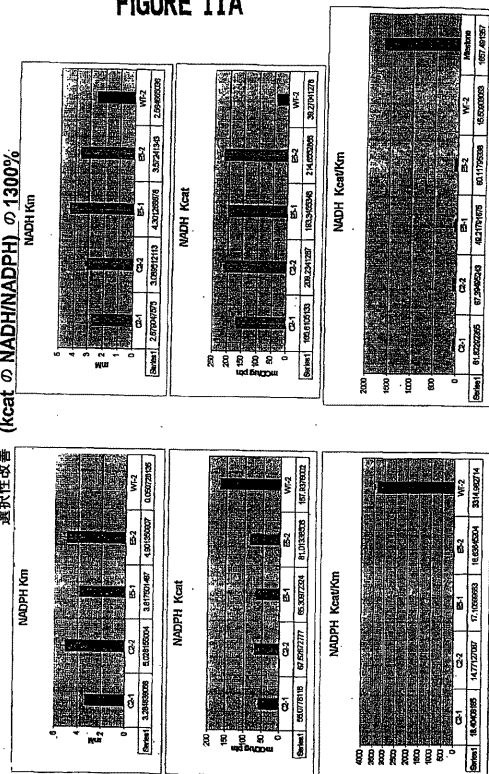
種々の濃度のシロアタマで突然変異体の当初反応速度



【図 1 1 A】

FIGURE 11A

2 PDA™ 設計 & WT TR の動力学パラメーター
変化コファクター-特異性 & 改善触媒効率
選択性改善 (kcat の NADH/NADPH) の 1300%



C2 は M-RYN 及び E5 は L-RYN, WT は I-RRR

【 図 1 2 】

TR-2 設計からの最良変異

Sp. Act. 1.2 mM	NADPH	NADH
RYN	25	106
RFN	24	108
RYN-A	19	97
RFN-A	15	81
WT-RRR	320	24
REN,RLN,RRN	65, 70, 340	65-75

FIGURE 12

【 図 1 5 】

廣範&改善觸媒活性

名称	実験 #1 (2 変数の平均)				実験 #2 (4 変数の平均)				動力学パラメーター			
	10mM NADPH	1.2mM NADPH	30mM NADH	1.2mM NADH	10mM NADPH	1.2mM NADPH	30mM NADH	1.2mM NADH	K _M	K _{cat}	K _{cat/K_M}	単位
R4W フラット ヒット	WR7	30.5.21	32.5.26	105.16	35.2.21	397.01	21.12.63	21.12.63	35.2.21	318.17	22.9.13	
	WLS	13.35	30.50	74.51	12.25.53	41.62	23.53.51	23.53.51	2.17	413.04	190.08	
	WHS	18.53	32.71	65.60	118.21	38.16	121.51	121.51	2.44	358.32	162.49	
	WRS	32.73	318.02	61.73	69.40	376.58	123.26	123.26	1.47	257.17	174.71	
	WIS	45.08	87.76	83.01	125.57	111.75	181.53	181.53	2.35	225.55	105.78	
WXX フラット ヒット	WFC	0.88	1.21	60.72	23.53.51	2.19	30.12.42	30.12.42	2.28	225.55	222.43	
	WFR	476.20	569.27	72.03	23.53.51	731.63	161.03	161.03	2.27	225.55	92.91	
	WNG	15.32	31.20	72.03	116.76	37.67	130.09	130.09	2.21	388.73	174.83	
	WVG	58.72	107.29	49.83	92.51	134.90	110.87	110.87	3.04	358.45	120.50	
	RTN	8.19	14.27	34.91	57.12	15.92	68.01	68.01	4.79	189.45	94.48	
TR-2	RRR-WT	200.70	10.09	17.18	23.53.51	10.00			1.01	57.61	25.74	

【 図 1 6 B - 1 】

N 末端に S タグ、C 末端に His タグを有する、WMG 変異体コード配列
(5238-26)

1	tcgagccaca	ccaccaccac	cactgagatc	cgggtgctaa	caaaagccga	a
61	aggaagctg	agttggctgc	tgccaccogt	gagcaataac	tagcataacc	ccttggggcc
121	ctaaacggg	tcttgagggg	tttttgcgtg	aaaggaggaa	ctatatccgg	attggcgaat
181	ggacgcgoc	ctgtagcggc	gcattaagcg	cggcggggtg	ggtggttacg	cgcagcgtga
241	cgtctacat	tgccagcgcc	ctagcgccgc	ctcctttgcg	ttttctccct	tcctttctcg
301	cacgttgct	cggctttccc	cgtcaagctc	taaatcgggg	gctcccttta	gggttcogat
361	tagtgcttt	acggcaacct	gaccccaaaa	aacttgatta	gggtgatggt	tcaogtagtg
421	gccatcgcc	ctgatagaog	gtttttgcgc	ctttgacggt	ggagtocagc	ttctttaata
481	tggactctt	gttccaaaat	ggaacaacac	tcaaccttat	ctcggtctat	tcttttgatt
541	ataagggat	ttttgcgatt	tcggcctatt	ggttaaaaaa	tgagotgatt	taacaaaaat
601	taacgcgaa	ttttaacaaa	atattaacgt	ttacaatttc	aggtggcact	tttcggggaa
661	tgctgcggc	aaccctctat	tgtttttttt	tctaaataca	ttcaaatatg	tatccgctca
721	gaattaatt	cttagaaaaa	ctcatcgagc	atcaaatgaa	actgcaattt	attcatatca
781	gattataat	taccatattt	ttgaaaaaag	cgtttctgta	atgaaggaga	aaactcacgc
841	ggcagtttc	ataggatggc	aagatcctgg	tatcggtctg	ogatccogac	togtocaaca
901	caatacaca	ctattaattt	ccctctgtca	aaaataaggt	tatcaagtga	gaaatcacca
961	gagtgagca	ctgaatccgg	tgagaatggc	aaaagtttat	gcattttctt	ccagacttgt
1021	caacaggcg	agccattaa	ctcgtcatca	aatcatcgct	cataaaccaa	acogttattc
1081	ttcgtgatt	ggcctcgagc	gagacgaat	acgcgatcgc	tgttaaaagg	acaattacaa
1141	caggaattc	aatgcaacgc	gcgcaggaa	actgccagcg	catcaacaat	attttcacct
1201	aatcaggat	attcttctaa	tacctggaat	gctgttttcc	cggggatcgc	agtggtgagt
1261	accatgatc					

【図 1 6 A - 1】

N 末端に S タグ、C 末端に His タグを有する、WVR 変異体コード配列
(5238-26)

```

1      tcgagcacca cccaccacac cactgagatc cggctgctaa caaagccoga a
aggaaactg
61     agttggctgc tgcacccctg gagcaataac tagcataacc ccttggggcc t
ctaaacggg
121    tcttgagggg ttttttgcgt aaaggaggaa ctatatccgg attggcgaat g
ggacgcggg
181    ctgtagcggc gcattaaagg cggcggggtg gtggtgtaac cgaagcgtga c
cgtcacact
241    tgccagcgcc ctacgcggcc ctcccttgcg ttctctccct tccctttctg c
cacgtttgc
301    cggcctttcc cgtcaagctc taactcgggg gctcccttta gggttccgat t
tagtgcctt
361    acggcacctc gaccocaaaa aacttgatta gggtagtggt tcacgtatgt g
gccatcgcc
421    ctgatagacg gtttttcgcc ctttgcagtt ggagtcacag ttctttaata g
tggactctc
481    gttccaaact ggaacaacac tcaaccctat ctcggtctat tcttttgatt t
ataagggat
541    ttgtgcgatt tcggcctatt gggtaaaaaa tgagctgatt taacaaaaat t
taacgcgaa
601    tttaacaaa atattaacgt ttacaatttc aggtggcact tttcggggaa a
tgtgcggcg
661    aacctccatt tgtttatttt tctaataata ttcaaatatg tatccggtca t
gaattaatt
721    cttagaaaaa ctctcagcgc atcaaatgaa actgcgaatt attcatatca t
gattatcaa
781    taccatattt ttgaaaaagg ctggttcgtg atgaaggaga aaactcaacc g
ggcagtttc
841    ataggatggc aagatccctg tatcgttcgt cagttccgac tgcctcaaca t
caatacaac
901    ctattaattt ccctctctca aaaaataagg ttcaaatgta gaataacca t
gagtgcaga
961    ctgaatccgc tgagaatgcg aaaagtttat gcattttctt ccagacttgt t
caacaggcc
1021   agccattacg ctctgcatac aaatcactcg catcaaccaa accgttatto a
tctgtgatt
1081   gcgcctgagc gagacgaat accgcgatgc tgttaaaagg acaattaca a
aatgcagag
1141   aatgcacccg gcgcaggaac atgccacgcy catcaacaat attttcaact g
aatcagagt
1201   attcttctaa taactggaat gctgttttcc cggggatcgc agtggtgagt a
acctgcgat

```

FIGURE 16A—1

【図 1 6 C - 1】

N 末端に S タグ、C 末端に His タグを有する、WIS 変異体コード配列
(5238-26)

a	ggaagaagct	1	tcgagaccaca	ccaccacccaac	cactgagatc	cgggtgctaa	caaagcccga	a
c	61	agttagctgc	tgccaccogct	gagcaataac	tagcataaac	cottggggcc	t	
c	ctaacycggt	121	tcttgaggggg	tttttttgcgt	aaaggaggaaa	ctatatccgg	attggcgaat	g
g	ggacgcggcg	181	ctgtagcggcg	gcattaagaag	cggcggggtg	gggtggttaag	cgcagcgtga	c
c	cgcatacact	241	tgccagcggcc	ctagcgcgcg	ctcctttcgc	tttcttcocct	tctctttctgc	c
c	caogtltcgc	301	cggcgtttccc	cgctcaagctc	taaatcgggg	gctcccttta	gggttcogcat	t
t	tagtgtcttt	361	acggcacctc	gacccccaaaa	aaccttgatta	gggtgatggt	tcacgtatgtg	g
g	gcocatcgcc	421	ctgatagacg	gttttttcgcg	ctttgacgnt	ggagtcacag	ttctttaata	g
t	tggactcttt	481	gttccaaact	ggaacaacac	tcAACccat	ctcggcttat	tcttttgatt	t
t	ataaggagat	541	ttttgccgatt	tgcggcttatt	ggttaaaaaa	tgagctgatt	taacaaaaat	t
t	taacgcgaa	601	ttttaacaaa	atattaaactg	ttacaatttc	aggtggcact	tttcggggaa	a
a	tgtgcggcg	661	aaacocctatt	tgtttatttt	tctaaatttc	ttcaaatatg	tatccgctca	t
t	gaattaatt	721	cttagaaaaa	ctcactcgagc	atcaaatgaa	actgcaattt	attcatatca	g
g	gattatcaaa	781	tacoattatt	ttgaaaaagc	cgtttctctga	atgaaggaga	aaactcaacc	a
a	ggcagttctg	841	ataggatggc	aagatcctgg	tatcggctcg	cgattccgac	tctgtcaaca	t
t	caatacaac	901	ctattaattt	ccctctctga	aaaataaggt	tatcaagtga	gaaatcacca	t
t	gagtgcaga	961	ctgaatccgg	tgagaatggc	aaaagtttat	gcattttctt	ccagactttgt	t
t	caacagccg	1021	agccatttacg	ctcgtcatca	aaatcacctg	catcaacca	acogttatto	a
a	ttctgctgct	1081	gcgcctgagc	gagacgaat	acgcgatcgc	tgttaaaagg	acaattacaa	a
g	caagcaatgc	1141	aatgcaacccg	gcgcaggaac	actgcacagc	catcaacaat	attttcaact	g
a	aaccagatg	1201	attctottaa	taoctggaat	gotgttttcc	cggggatcgc	agtggtagt	a
a	accatcatg							

FIGURE 16C – 1

FIGURE 16B — 1

【図 16D-1】

N 末端に S タグ、C 末端に His タグを有する、WMS 変異体コード配列
(5238-26)

```

1 tcgagcacca ccaccaccac cactgagatc cggctgctaa caaagcccca a
aggaagctg 61 agttggctgc tgccaccgct gagcaataac tagcataacc ccttggggcc t
ctaaacggg 121 tcttgagggg ttttttgctg aaaggaggaa ctatatccgg attggcgaaat g
ggacgcgcc 181 ctgtagcggc gcattaagcg cggcgggtgt ggtgggttacg cgcagcgtga c
cgctacact 241 tgccagcgcc ctagcgcccg ctccctttgcg tttcttccct tcttttctcg c
cacgtttcgc 301 cggctttccc cgtcaagctc taaatcgggg gctcccttta gggttccgat t
tagtgcttt 361 acggcacctc gaccccaaaa aacttgatta gggtagtggt tcacgtagtg g
gccatcgcc 421 ctgatatagc gtttttgcgc ctttgacggt ggagtccacg ttctttaata g
tggactctt 481 gttccaaact ggaacaacac tcaacctat ctoggtctat tcttttgatt t
ataagggat 541 tttgccgatt tcggcctatt ggttaaaaaa tgagctgatt taacaaaaat t
taacgcgaa 601 ttttaacaaa atattaacgt ttacaatttc aggtggcact tttcggggaa a
tgtgcgcgg 661 aacccctatt tgtttatttt tctaaataca ttcaaatatg tatccgctca t
gaattaatt 721 cttagaaaaa ctcatcgagc atcaaatgaa actgcaattt attcatatca g
gattatcaa 781 taccatattt ttgaaaaagc cgtttctgta atgaaggaga aaactcacgc a
ggcagttcc 841 ataggatggc aagatcctgg tatcggtctg cgattccgac tegtccaaca t
caatacaac 901 ctattaattt ccctcgtca aaaataaggt tatcaagtga gaaatcacca t
jagtgacga 961 ctgaatccgg tgagaatggc aaaagtttat gcatttcttt ccagacttgt t
caacaggcc 1021 agccattacg ctcgctatca aaatcaactg catcaaccaa accgttattc a
ttcgtgatt 1081 gcgcctgagc gagaagaaat acgcgatcgc tgttaaaagg acaattacaa a
caggaatcg 1141 aatgcaaccg gcgcaggaaac actgccagcg catcaacaat attttcacct g
aatcaggat 1201 attcttctaa tacctggaat gctgttttcc cggggatcgc agtggtagt a
accatgcat 1261 catcaggagt acggataaaa tgcttgatgg tcggaagagg cataaattcc g

```

FIGURE 16D - 1

【図 16E-1】

N 末端に S タグ、C 末端に His タグを有する、WLS 変異体コード配列
(5238-26)

```

1 tcgagcacca ccaccaccac cactgagatc cggctgctaa caaagcccca a
aggaagctg 61 agttggctgc tgccaccgct gagcaataac tagcataacc ccttggggcc t
ctaaacggg 121 tcttgagggg ttttttgctg aaaggaggaa ctatatccgg attggcgaaat g
ggacgcgcc 181 ctgtagcggc gcattaagcg cggcgggtgt ggtgggttacg cgcagcgtga c
cgctacact 241 tgccagcgcc ctagcgcccg ctccctttgcg tttcttccct tcttttctcg c
cacgtttcgc 301 cggctttccc cgtcaagctc taaatcgggg gctcccttta gggttccgat t
tagtgcttt 361 acggcacctc gaccccaaaa aacttgatta gggtagtggt tcacgtagtg g
gccatcgcc 421 ctgatatagc gtttttgcgc ctttgacggt ggagtccacg ttctttaata g
tggactctt 481 gttccaaact ggaacaacac tcaacctat ctoggtctat tcttttgatt t
ataagggat 541 tttgccgatt tcggcctatt ggttaaaaaa tgagctgatt taacaaaaat t
taacgcgaa 601 ttttaacaaa atattaacgt ttacaatttc aggtggcact tttcggggaa a
tgtgcgcgg 661 aacccctatt tgtttatttt tctaaataca ttcaaatatg tatccgctca t
gaattaatt 721 cttagaaaaa ctcatcgagc atcaaatgaa actgcaattt attcatatca g
gattatcaa 781 taccatattt ttgaaaaagc cgtttctgta atgaaggaga aaactcacgc a
ggcagttcc 841 ataggatggc aagatcctgg tatcggtctg cgattccgac tegtccaaca t
caatacaac 901 ctattaattt ccctcgtca aaaataaggt tatcaagtga gaaatcacca t
jagtgacga 961 ctgaatccgg tgagaatggc aaaagtttat gcatttcttt ccagacttgt t
caacaggcc 1021 agccattacg ctcgctatca aaatcaactg catcaaccaa accgttattc a
ttcgtgatt 1081 gcgcctgagc gagaagaaat acgcgatcgc tgttaaaagg acaattacaa a
caggaatcg 1141 aatgcaaccg gcgcaggaaac actgccagcg catcaacaat attttcacct g
aatcaggat 1201 attcttctaa tacctggaat gctgttttcc cggggatcgc agtggtagt a
accatgcat 1261 catcaggagt acggataaaa tgcttgatgg tcggaagagg cataaattcc g

```

FIGURE 16E - 1

【図 16F-1】

N 末端に S タグ、C 末端に His タグを有する、WRT 変異体コード配列
(5238-26)

```

1 tcgagcacca ccaccaccac cactgagatc cggctgctaa caaagcccca a
aggaagctg 61 agttggctgc tgccaccgct gagcaataac tagcataacc ccttggggcc t
ctaaacggg 121 tcttgagggg ttttttgctg aaaggaggaa ctatatccgg attggcgaaat g
ggacgcgcc 181 ctgtagcggc gcattaagcg cggcgggtgt ggtgggttacg cgcagcgtga c
cgctacact 241 tgccagcgcc ctagcgcccg ctccctttgcg tttcttccct tcttttctcg c
cacgtttcgc 301 cggctttccc cgtcaagctc taaatcgggg gctcccttta gggttccgat t
tagtgcttt 361 acggcacctc gaccccaaaa aacttgatta gggtagtggt tcacgtagtg g
gccatcgcc 421 ctgatatagc gtttttgcgc ctttgacggt ggagtccacg ttctttaata g
tggactctt 481 gttccaaact ggaacaacac tcaacctat ctoggtctat tcttttgatt t
ataagggat 541 tttgccgatt tcggcctatt ggttaaaaaa tgagctgatt taacaaaaat t
taacgcgaa 601 ttttaacaaa atattaacgt ttacaatttc aggtggcact tttcggggaa a
tgtgcgcgg 661 aacccctatt tgtttatttt tctaaataca ttcaaatatg tatccgctca t
gaattaatt 721 cttagaaaaa ctcatcgagc atcaaatgaa actgcaattt attcatatca g
gattatcaa 781 taccatattt ttgaaaaagc cgtttctgta atgaaggaga aaactcacgc a
ggcagttcc 841 ataggatggc aagatcctgg tatcggtctg cgattccgac tegtccaaca t
caatacaac 901 ctattaattt ccctcgtca aaaataaggt tatcaagtga gaaatcacca t
jagtgacga 961 ctgaatccgg tgagaatggc aaaagtttat gcatttcttt ccagacttgt t
caacaggcc 1021 agccattacg ctcgctatca aaatcaactg catcaaccaa accgttattc a
ttcgtgatt 1081 gcgcctgagc gagaagaaat acgcgatcgc tgttaaaagg acaattacaa a
caggaatcg 1141 aatgcaaccg gcgcaggaaac actgccagcg catcaacaat attttcacct g
aatcaggat 1201 attcttctaa tacctggaat gctgttttcc cggggatcgc agtggtagt a
accatgcat 1261 catcaggagt acggataaaa tgcttgatgg tcggaagagg cataaattcc g

```

FIGURE 16F - 1

【図 16G-1】

N 末端に S タグ、C 末端に His タグを有する、RYN 変異体コード配列
(5238-26)

```

1 tcgagcacca ccaccaccac cactgagatc cggctgctaa caaagcccca a
aggaagctg 61 agttggctgc tgccaccgct gagcaataac tagcataacc ccttggggcc t
ctaaacggg 121 tcttgagggg ttttttgctg aaaggaggaa ctatatccgg attggcgaaat g
ggacgcgcc 181 ctgtagcggc gcattaagcg cggcgggtgt ggtgggttacg cgcagcgtga c
cgctacact 241 tgccagcgcc ctagcgcccg ctccctttgcg tttcttccct tcttttctcg c
cacgtttcgc 301 cggctttccc cgtcaagctc taaatcgggg gctcccttta gggttccgat t
tagtgcttt 361 acggcacctc gaccccaaaa aacttgatta gggtagtggt tcacgtagtg g
gccatcgcc 421 ctgatatagc gtttttgcgc ctttgacggt ggagtccacg ttctttaata g
tggactctt 481 gttccaaact ggaacaacac tcaacctat ctoggtctat tcttttgatt t
ataagggat 541 tttgccgatt tcggcctatt ggttaaaaaa tgagctgatt taacaaaaat t
taacgcgaa 601 ttttaacaaa atattaacgt ttacaatttc aggtggcact tttcggggaa a
tgtgcgcgg 661 aacccctatt tgtttatttt tctaaataca ttcaaatatg tatccgctca t
gaattaatt 721 cttagaaaaa ctcatcgagc atcaaatgaa actgcaattt attcatatca g
gattatcaa 781 taccatattt ttgaaaaagc cgtttctgta atgaaggaga aaactcacgc a
ggcagttcc 841 ataggatggc aagatcctgg tatcggtctg cgattccgac tegtccaaca t
caatacaac 901 ctattaattt ccctcgtca aaaataaggt tatcaagtga gaaatcacca t
jagtgacga 961 ctgaatccgg tgagaatggc aaaagtttat gcatttcttt ccagacttgt t
caacaggcc 1021 agccattacg ctcgctatca aaatcaactg catcaaccaa accgttattc a
ttcgtgatt 1081 gcgcctgagc gagaagaaat acgcgatcgc tgttaaaagg acaattacaa a
caggaatcg 1141 aatgcaaccg gcgcaggaaac actgccagcg catcaacaat attttcacct g
aatcaggat 1201 attcttctaa tacctggaat gctgttttcc cggggatcgc agtggtagt a
accatgcat 1261 catcaggagt acggataaaa tgcttgatgg tcggaagagg cataaattcc g

```

FIGURE 16G - 1

【図 16H-1】

N 末端に S タグ、C 末端に His タグを有する、RFN-A 変異体コード配列
(5238-26)

```

1 tcgagcacca ccaccaccac cactgagatc cggctgctaa caaagcccca a
aggaagctg
61 agttggctgc tgccaccgct gagcaataac tagcataacc ccttggggcc t
ctaaacggg
121 tcttgagggg ttttttgctg aaaggaggaa ctatatccgg attggcgaat g
ggacgcggcc
181 ctgtagcggc gcattaagcg cggcgggtgt ggtggttacg cgcagcgtga c
cgctacact
241 tgccagcgcc cttagcgccc ctcttttcgc tttcttccct tcctttctcg c
cacgtttcgc
301 cggctttccc cgtcaagctc taaatcgggg gctcccttta gggttccgat t
tagtgcttt
361 acggcacctc gaccccaaaa aacttgatta gggatgaggt tcacgtagt g
gccatcgcc
421 ctgatagacg gtttttcgoc ctttgacgtt ggagtccacg ttctttaata g
tggactctt
481 gttccaaact ggaacaacac tcaaccctat ctcggtctat tcttttgatt t
ataagggat
541 tttgcccatt tcggcctatt ggttaaaaaa tgagctgatt taacaaaaat t
taacgcgaa
601 ttttaacaaa atattaacgt ttacaatttc aggtggcact tttcggggaa a
tgtgcgcgg
661 aacccctatt tgtttatttt tctaaatata ttcaaatatg tatccgctca t
gaattaatt
721 cttagaaaaa ctcatcgagc atcaaatgaa actgcaattt attcatatca g
gattatcaa
781 taccatattt ttgaaaaagc cgtttctgta atgaaggaga aaactcacgc a
ggcagttcc
841 ataggatggc aagatcctgg tatcggtctg cgattccgac tcgtccaaca t
caatacaac
901 ctattaattt cccctcgta aaaataaggt tatcaagtga gaaatcacca t
gagtgcgca
961 ctgaatccgg tgagaatggc aaaagtttat gcatttcttt ccagacttgt t
caacagggc
1021 agccattacg ctcgctatca aaatcaactg catcaaccaa accgttattc a
ttcgtgatt
1081 gcgcctgagc gagacgaaat acgcgatcgc tgttaaaagg acaattacaa a
caggaatcg
1141 aatgcaacgg gcgcaggaaac actgccagcg catcaacaat attttcacct g
aatcaggat
1201 attcttctaa tacctggaat gctgttttcc cggggatcgc agtggtgagt a
accatgcat

```

FIGURE 16H-1

【図 16I-1】

N 末端に S タグ、C 末端に His タグを有する、RFN 変異体コード配列
(5238-26)

```

1 tcgagcacca ccaccaccac cactgagatc cggctgctaa caaagcccca a
aggaagctg
61 agttggctgc tgccaccgct gagcaataac tagcataacc ccttggggcc t
ctaaacggg
121 tcttgagggg ttttttgctg aaaggaggaa ctatatccgg attggcgaat g
ggacgcggcc
181 ctgtagcggc gcattaagcg cggcgggtgt ggtggttacg cgcagcgtga c
cgctacact
241 tgccagcgcc cttagcgccc ctcttttcgc tttcttccct tcctttctcg c
cacgtttcgc
301 cggctttccc cgtcaagctc taaatcgggg gctcccttta gggttccgat t
tagtgcttt
361 acggcacctc gaccccaaaa aacttgatta gggatgaggt tcacgtagt g
gccatcgcc
421 ctgatagacg gtttttcgoc ctttgacgtt ggagtccacg ttctttaata g
tggactctt
481 gttccaaact ggaacaacac tcaaccctat ctcggtctat tcttttgatt t
ataagggat
541 tttgcccatt tcggcctatt ggttaaaaaa tgagctgatt taacaaaaat t
taacgcgaa
601 ttttaacaaa atattaacgt ttacaatttc aggtggcact tttcggggaa a
tgtgcgcgg
661 aacccctatt tgtttatttt tctaaatata ttcaaatatg tatccgctca t
gaattaatt
721 cttagaaaaa ctcatcgagc atcaaatgaa actgcaattt attcatatca g
gattatcaa
781 taccatattt ttgaaaaagc cgtttctgta atgaaggaga aaactcacgc a
ggcagttcc
841 ataggatggc aagatcctgg tatcggtctg cgattccgac tcgtccaaca t
caatacaac
901 ctattaattt cccctcgta aaaataaggt tatcaagtga gaaatcacca t
gagtgcgca
961 ctgaatccgg tgagaatggc aaaagtttat gcatttcttt ccagacttgt t
caacagggc
1021 agccattacg ctcgctatca aaatcaactg catcaaccaa accgttattc a
ttcgtgatt
1081 gcgcctgagc gagacgaaat acgcgatcgc tgttaaaagg acaattacaa a
caggaatcg
1141 aatgcaacgg gcgcaggaaac actgccagcg catcaacaat attttcacct g
aatcaggat
1201 attcttctaa tacctggaat gctgttttcc cggggatcgc agtggtgagt a
accatgcat

```

FIGURE 16I-1

【図 16J-1】

N 末端に S タグ、C 末端に His タグを有する、RRR-WT 変異体コード配列
(5238-26)

```

1 tcgagcacca ccaccaccac cactgagatc cggctgctaa caaagcccca a
aggaagctg
61 agttggctgc tgccaccgct gagcaataac tagcataacc ccttggggcc t
ctaaacggg
121 tcttgagggg ttttttgctg aaaggaggaa ctatatccgg attggcgaat g
ggacgcggcc
181 ctgtagcggc gcattaagcg cggcgggtgt ggtggttacg cgcagcgtga c
cgctacact
241 tgccagcgcc cttagcgccc ctcttttcgc tttcttccct tcctttctcg c
cacgtttcgc
301 cggctttccc cgtcaagctc taaatcgggg gctcccttta gggttccgat t
tagtgcttt
361 acggcacctc gaccccaaaa aacttgatta gggatgaggt tcacgtagt g
gccatcgcc
421 ctgatagacg gtttttcgoc ctttgacgtt ggagtccacg ttctttaata g
tggactctt
481 gttccaaact ggaacaacac tcaaccctat ctcggtctat tcttttgatt t
ataagggat
541 tttgcccatt tcggcctatt ggttaaaaaa tgagctgatt taacaaaaat t
taacgcgaa
601 ttttaacaaa atattaacgt ttacaatttc aggtggcact tttcggggaa a
tgtgcgcgg
661 aacccctatt tgtttatttt tctaaatata ttcaaatatg tatccgctca t
gaattaatt
721 cttagaaaaa ctcatcgagc atcaaatgaa actgcaattt attcatatca g
gattatcaa
781 taccatattt ttgaaaaagc cgtttctgta atgaaggaga aaactcacgc a
ggcagttcc
841 ataggatggc aagatcctgg tatcggtctg cgattccgac tcgtccaaca t
caatacaac
901 ctattaattt cccctcgta aaaataaggt tatcaagtga gaaatcacca t
gagtgcgca
961 ctgaatccgg tgagaatggc aaaagtttat gcatttcttt ccagacttgt t
caacagggc
1021 agccattacg ctcgctatca aaatcaactg catcaaccaa accgttattc a
ttcgtgatt
1081 gcgcctgagc gagacgaaat acgcgatcgc tgttaaaagg acaattacaa a
caggaatcg
1141 aatgcaacgg gcgcaggaaac actgccagcg catcaacaat attttcacct g
aatcaggat
1201 attcttctaa tacctggaat gctgttttcc cggggatcgc agtggtgagt a
accatgcat
1261 catcaggagt acggataaaa tgcttgatgg tcggaagagg cataaattcc g

```

FIGURE 16J-1

【図 16K-1】

N 末端に S タグ 有する、WVG 変異体コード配列
(5238 -6335)

```

1 tcgagcacca ccaccaccac cactgagatc cggctgctaa caaagcccca a
aggaagctg
61 agttggctgc tgccaccgct gagcaataac tagcataacc ccttggggcc t
ctaaacggg
121 tcttgagggg ttttttgctg aaaggaggaa ctatatccgg attggcgaat g
ggacgcggcc
181 ctgtagcggc gcattaagcg cggcgggtgt ggtggttacg cgcagcgtga c
cgctacact
241 tgccagcgcc cttagcgccc ctcttttcgc tttcttccct tcctttctcg c
cacgtttcgc
301 cggctttccc cgtcaagctc taaatcgggg gctcccttta gggttccgat t
tagtgcttt
361 acggcacctc gaccccaaaa aacttgatta gggatgaggt tcacgtagt g
gccatcgcc
421 ctgatagacg gtttttcgoc ctttgacgtt ggagtccacg ttctttaata g
tggactctt
481 gttccaaact ggaacaacac tcaaccctat ctcggtctat tcttttgatt t
ataagggat
541 tttgcccatt tcggcctatt ggttaaaaaa tgagctgatt taacaaaaat t
taacgcgaa
601 ttttaacaaa atattaacgt ttacaatttc aggtggcact tttcggggaa a
tgtgcgcgg
661 aacccctatt tgtttatttt tctaaatata ttcaaatatg tatccgctca t
gaattaatt
721 cttagaaaaa ctcatcgagc atcaaatgaa actgcaattt attcatatca g
gattatcaa
781 taccatattt ttgaaaaagc cgtttctgta atgaaggaga aaactcacgc a
ggcagttcc
841 ataggatggc aagatcctgg tatcggtctg cgattccgac tcgtccaaca t
caatacaac
901 ctattaattt cccctcgta aaaataaggt tatcaagtga gaaatcacca t
gagtgcgca
961 ctgaatccgg tgagaatggc aaaagtttat gcatttcttt ccagacttgt t
caacagggc
1021 agccattacg ctcgctatca aaatcaactg catcaaccaa accgttattc a
ttcgtgatt
1081 gcgcctgagc gagacgaaat acgcgatcgc tgttaaaagg acaattacaa a
caggaatcg
1141 aatgcaacgg gcgcaggaaac actgccagcg catcaacaat attttcacct g
aatcaggat
1201 attcttctaa tacctggaat gctgttttcc cggggatcgc agtggtgagt a
accatgcat
1261 catcaggagt acggataaaa tgcttgatgg tcggaagagg cataaattcc g

```

FIGURE 16K-1

【図 16 L - 1】

N末端にSタグを有する、WRS 変異体コード配列
(5238 -6335)

```

1 tcgagcacca ccaccaccac cactgagatc cggctgctaa caaagccoga a
aggaagctg
61 agttggctgc tgccaccgct gagcaataac tagcataacc ccttggggcc t
ctaaacggg
121 tottgagggg ttttttgcgt aaaggaggaa ctatatccgg attggcgaat g
ggacgcggcc
181 ctgtagcggc gcattaagcg cggcgggtgt ggtggttacg cgcagcgtga c
cgctacact
241 tgccagcgcc cttagcgccg ctcccttgcg tttcttccct tctttctcgc c
cacgtttgc
301 cggctttccc cgtcaagctc taaatcgggg gctcccttta gggttccgat t
tagtgcttt
361 acggcacctc gaccccaaaa aacttgatta gggtagtggt tcacgtagt g
gccatcgcc
421 ctgataagcg gtttttgcgc ctttgacgtt ggagtccacg ttctttaata g
tggaactctt
481 gttccaaact ggaacaacac tcaacctat ctoggtctat tcttttgatt t
ataagggat
541 tttgcccatt tcggcctatt ggttaaaaaa tgagctgatt taacaaaaat t
taacgcgaa
601 ttttaacaaa atattaacgt ttacaatttc aggtggcact tttcggggaa a
tgtgcggcg
661 aacccctatt tgtttatttt tctaaataca ttcaaatatg tatccgctca t
gaattaatt
721 cttagaaaaa ctcatcgagc atcaaatgaa actgcaattt attcatatca g
gattatcaa
781 taccatattt ttgaaaaagc cgtttctgta atgaaggaga aaactcaccg a
ggcagttcc
841 ataggatggc aagatcctcg tatcggtctg cgaatccgac tcgtccaaca t
caatacaac
901 ctattaattt cccctcgtca aaaataaggt tatcaagtga gaaatcacca t
gagtgacga
961 ctgaatccgg tgagaatggc aaaagtttat gcattttctt ccagacttgt t
caacaggcc
1021 agccattacg ctogtcatca aaatcaactg catcaaccaa accgttatcc a
ttcgtgatt
1081 gcgcctgagc gagacgaaat acgcgacgcg tgttaaaagg acaattacaa a
caggaatcg
1141 aatgcaacgg gcgcaggaaac actgccagcg catcaacaat attttcacct g
aatcaggat
1201 attcttctaa tacctggaat gctgttttcc cggggatcgc agtggtgagt a
accatgcat
1261

```

FIGURE 16L - 1

【図 16 M - 1】

N末端にSタグ、WFO 変異体コード配列
(5236 -6335)

```

1 tcgagcacca ccaccaccac cactgagatc cggctgctaa caaagccoga a
aggaagctg
61 agttggctgc tgccaccgct gagcaataac tagcataacc ccttggggcc t
ctaaacggg
121 tottgagggg ttttttgcgt aaaggaggaa ctatatccgg attggcgaat g
ggacgcggcc
181 ctgtagcggc gcattaagcg cggcgggtgt ggtggttacg cgcagcgtga c
cgctacact
241 tgccagcgcc cttagcgccg ctcccttgcg tttcttccct tctttctcgc c
cacgtttgc
301 cggctttccc cgtcaagctc taaatcgggg gctcccttta gggttccgat t
tagtgcttt
361 acggcacctc gaccccaaaa aacttgatta gggtagtggt tcacgtagt g
gccatcgcc
421 ctgataagcg gtttttgcgc ctttgacgtt ggagtccacg ttctttaata g
tggaactctt
481 gttccaaact ggaacaacac tcaacctat ctoggtctat tcttttgatt t
ataagggat
541 tttgcccatt tcggcctatt ggttaaaaaa tgagctgatt taacaaaaat t
taacgcgaa
601 ttttaacaaa atattaacgt ttacaatttc aggtggcact tttcggggaa a
tgtgcggcg
661 aacccctatt tgtttatttt tctaaataca ttcaaatatg tatccgctca t
gaattaatt
721 cttagaaaaa ctcatcgagc atcaaatgaa actgcaattt attcatatca g
gattatcaa
781 taccatattt ttgaaaaagc cgtttctgta atgaaggaga aaactcaccg a
ggcagttcc
841 ataggatggc aagatcctcg tatcggtctg cgaatccgac tcgtccaaca t
caatacaac
901 ctattaattt cccctcgtca aaaataaggt tatcaagtga gaaatcacca t
gagtgacga
961 ctgaatccgg tgagaatggc aaaagtttat gcattttctt ccagacttgt t
caacaggcc
1021 agccattacg ctogtcatca aaatcaactg catcaaccaa accgttatcc a
ttcgtgatt
1081 gcgcctgagc gagacgaaat acgcgacgcg tgttaaaagg acaattacaa a
caggaatcg
1141 aatgcaacgg gcgcaggaaac actgccagcg catcaacaat attttcacct g
aatcaggat
1201 attcttctaa tacctggaat gctgttttcc cggggatcgc agtggtgagt a
accatgcat
1261

```

FIGURE 16M - 1

【図 16 N - 1】

N末端にSタグを有する、NTR-WT (RRR-WT) コード配列
(5238-6335)

```

1 tcgagcacca ccaccaccac cactgagatc cggctgctaa caaagccoga a
aggaagctg
61 agttggctgc tgccaccgct gagcaataac tagcataacc ccttggggcc t
ctaaacggg
121 tottgagggg ttttttgcgt aaaggaggaa ctatatccgg attggcgaat g
ggacgcggcc
181 ctgtagcggc gcattaagcg cggcgggtgt ggtggttacg cgcagcgtga c
cgctacact
241 tgccagcgcc cttagcgccg ctcccttgcg tttcttccct tctttctcgc c
cacgtttgc
301 cggctttccc cgtcaagctc taaatcgggg gctcccttta gggttccgat t
tagtgcttt
361 acggcacctc gaccccaaaa aacttgatta gggtagtggt tcacgtagt g
gccatcgcc
421 ctgataagcg gtttttgcgc ctttgacgtt ggagtccacg ttctttaata g
tggaactctt
481 gttccaaact ggaacaacac tcaacctat ctoggtctat tcttttgatt t
ataagggat
541 tttgcccatt tcggcctatt ggttaaaaaa tgagctgatt taacaaaaat t
taacgcgaa
601 ttttaacaaa atattaacgt ttacaatttc aggtggcact tttcggggaa a
tgtgcggcg
661 aacccctatt tgtttatttt tctaaataca ttcaaatatg tatccgctca t
gaattaatt
721 cttagaaaaa ctcatcgagc atcaaatgaa actgcaattt attcatatca g
gattatcaa
781 taccatattt ttgaaaaagc cgtttctgta atgaaggaga aaactcaccg a
ggcagttcc
841 ataggatggc aagatcctcg tatcggtctg cgaatccgac tcgtccaaca t
caatacaac
901 ctattaattt cccctcgtca aaaataaggt tatcaagtga gaaatcacca t
gagtgacga
961 ctgaatccgg tgagaatggc aaaagtttat gcattttctt ccagacttgt t
caacaggcc
1021 agccattacg ctogtcatca aaatcaactg catcaaccaa accgttatcc a
ttcgtgatt
1081 gcgcctgagc gagacgaaat acgcgacgcg tgttaaaagg acaattacaa a
caggaatcg
1141 aatgcaacgg gcgcaggaaac actgccagcg catcaacaat attttcacct g
aatcaggat
1201 attcttctaa tacctggaat gctgttttcc cggggatcgc agtggtgagt a
accatgcat
1261 catcaggagt acggataaaa tgcttgatgg tcggaagagg cataaattcc g

```

FIGURE 16 N - 1

【図 16 O - 1】

N末端にSタグを有するRYN-M変異体コード配列
(5238-6335)

```

1 tcgagcacca ccaccaccac cactgagatc cggctgctaa caaagccoga a
aggaagctg
61 agttggctgc tgccaccgct gagcaataac tagcataacc ccttggggcc t
ctaaacggg
121 tottgagggg ttttttgcgt aaaggaggaa ctatatccgg attggcgaat g
ggacgcggcc
181 ctgtagcggc gcattaagcg cggcgggtgt ggtggttacg cgcagcgtga c
cgctacact
241 tgccagcgcc cttagcgccg ctcccttgcg tttcttccct tctttctcgc c
cacgtttgc
301 cggctttccc cgtcaagctc taaatcgggg gctcccttta gggttccgat t
tagtgcttt
361 acggcacctc gaccccaaaa aacttgatta gggtagtggt tcacgtagt g
gccatcgcc
421 ctgataagcg gtttttgcgc ctttgacgtt ggagtccacg ttctttaata g
tggaactctt
481 gttccaaact ggaacaacac tcaacctat ctoggtctat tcttttgatt t
ataagggat
541 tttgcccatt tcggcctatt ggttaaaaaa tgagctgatt taacaaaaat t
taacgcgaa
601 ttttaacaaa atattaacgt ttacaatttc aggtggcact tttcggggaa a
tgtgcggcg
661 aacccctatt tgtttatttt tctaaataca ttcaaatatg tatccgctca t
gaattaatt
721 cttagaaaaa ctcatcgagc atcaaatgaa actgcaattt attcatatca g
gattatcaa
781 taccatattt ttgaaaaagc cgtttctgta atgaaggaga aaactcaccg a
ggcagttcc
841 ataggatggc aagatcctcg tatcggtctg cgaatccgac tcgtccaaca t
caatacaac
901 ctattaattt cccctcgtca aaaataaggt tatcaagtga gaaatcacca t
gagtgacga
961 ctgaatccgg tgagaatggc aaaagtttat gcattttctt ccagacttgt t
caacaggcc
1021 agccattacg ctogtcatca aaatcaactg catcaaccaa accgttatcc a
ttcgtgatt
1081 gcgcctgagc gagacgaaat acgcgacgcg tgttaaaagg acaattacaa a
caggaatcg
1141 aatgcaacgg gcgcaggaaac actgccagcg catcaacaat attttcacct g
aatcaggat
1201 attcttctaa tacctggaat gctgttttcc cggggatcgc agtggtgagt a
accatgcat
1261 catcaggagt acggataaaa tgcttgatgg tcggaagagg cataaattcc g

```

FIGURE 16 O - 1

【図 16 P-1】

N末端にSタグを有するRYN-L変異体コード配列 (52 38-6335)

```

1 tcgagcacca ccaccaccac cactgagatc cggtgctaa caaagcccca a
aggaagctgt
61 agttggctgc tgcacccgct gagcaataac tagcataacc cctgggggoc t
ctaaacggcg
121 tcttgagggg ttttttgctg aaaggaggaa ctatatccgg attggcgaat g
ggacgcgcc
181 ctgtagcgcg gcattaagcg cggcggtgtt ggtggttacg cgcagcgtga c
cgctacaact
241 tgccagcgcc ctacgcggcg ctcccttgcg tttcttccct tcttctctcg c
cacgttgcg
301 cggttttccc cgtcaagctc taaatcgggg gtcctcctta gggttcgat t
tagtgcttt
361 acggcacctc gaccccaaaa aacttgatta ggtgtatggt tcacgtatgt g
gccatcgcc
421 ctgatagacg gtttttgcgc ctttgacggtt ggagtccaag ttctttaata g
tggactctt
481 gttccaaact ggaacaacac tcaaccccat ctcggtctat tcttttgatt t
ataagggat
541 tttgccgatt tcggcctatt ggttaaaaaa tgagctgatt taacaaaaat t
taacgcgaa
601 ttttaacaaa atattaacgt ttacaatttc aggtggcact ttcggggaaa a
tgtgcgcgg
661 aacccttatt tgtttttttt tctaaatata ttcaaatatg tatccgctca t
gaattaatt
721 cttagaaaaa ctcatcgagc atcaaatgaa actgcaattt attcatatca g
gattatcaa
781 taccatattt ttgaaaagcg cgtttctgtg atgaaggaga aaactcacg a
ggcagttcg
841 ataggatggc aagatcctcg tatcggtctg cgattccgac tcgtccaaca t
caatacaac
901 ctattaattt cccctcgctc aaaataaggt tatcaagtga gaaatcacca t
gagtgcaga
961 ctgaatcccg tgagaatggc aaaagtttat gcatttcttt ccagactgtt t
caacaggcc
1021 agccattacg ctcgctcatca aatcactcgc catcaaccaa accgtttatc a
ttcgtagatt
1081 ggcctcgagc gagcaagaat acgcgcatgc tgttaaaagg acaattacaa a
caggaaatcg
1141 aatgcaacgg gcgcaggaaac actgccagcg catcaacaat atttcaact g
aatcaggat
1201 attcttctaa taactcgaaat gctgttttcc cgggtagtgc agtgggtgag t
accatcgat
1261 catcaggagt acggataaaa tgcttagtagg tcgggaagag cataaattcc g

```

FIGURE 16P -1

【図 17 B】

使用された発現ベクターによってコードされるような、NTR WTおよび変異体タンパク質配列のアラインメント。突然変異は、青/太字で示す。Sタグ戦略で使用されたトロロビンのための切断部位は、下線を付し、赤矢によって示す。

FileUp
MSP: 365 型: P 子17: 750 ..

Name	Len	Check	Weight
RRR-WT-S-Tagged Protein	365	37	0
RYN-I-S-Tagged Protein	365	177	0
RYN-L-S-Tagged Protein	365	255	0
RYN-M-S-Tagged Protein	365	281	0

//

```

1
RRR-WT-S-Tagged Protein MKETAAAKE RQMDSDLG TLVPRGSMAD IRNGLESTN TRLCIVGSSP 50
RYN-I-S-Tagged Protein MKETAAAKE RQMDSDLG TLVPRGSMAD IRNGLESTN TRLCIVGSSP
RYN-L-S-Tagged Protein MKETAAAKE RQMDSDLG TLVPRGSMAD IRNGLESTN TRLCIVGSSP
RYN-M-S-Tagged Protein MKETAAAKE RQMDSDLG TLVPRGSMAD IRNGLESTN TRLCIVGSSP
51
RRR-WT-S-Tagged Protein AAHTAAIYAA RAELEPLIFE GWANDIAPG GOLTTTIDE NFGPFPEGIL 100
RYN-I-S-Tagged Protein AAHTAAIYAA RAELEPLIFE GWANDIAPG GOLTTTIDE NFGPFPEGIL
RYN-L-S-Tagged Protein AAHTAAIYAA RAELEPLIFE GWANDIAPG GOLTTTIDE NFGPFPEGIL
RYN-M-S-Tagged Protein AAHTAAIYAA RAELEPLIFE GWANDIAPG GOLTTTIDE NFGPFPEGIL
101
RRR-WT-S-Tagged Protein GVELTDKFRK QSERPTTIF TETVTKVDFS SKPKLFIDS KAILADAVIL 150
RYN-I-S-Tagged Protein GVELTDKFRK QSERPTTIF TETVTKVDFS SKPKLFIDS KAILADAVIL
RYN-L-S-Tagged Protein GVELTDKFRK QSERPTTIF TETVTKVDFS SKPKLFIDS KAILADAVIL
RYN-M-S-Tagged Protein GVELTDKFRK QSERPTTIF TETVTKVDFS SKPKLFIDS KAILADAVIL
151
RRR-WT-S-Tagged Protein ATGAVAKRLS FVGSBGSOG FWRNGISACA VCDGAPFIR NKPLAVIGGG 200
RYN-I-S-Tagged Protein ATGAVAKRLS FVGSBGSOG FWRNGISACA VCDGAPFIR NKPLAVIGGG
RYN-L-S-Tagged Protein ATGAVAKRLS FVGSBGSOG FWRNGISACA VCDGAPFIR NKPLAVIGGG
RYN-M-S-Tagged Protein ATGAVAKRLS FVGSBGSOG FWRNGISACA VCDGAPFIR NKPLAVIGGG
201
RRR-WT-S-Tagged Protein DSAMEEANFL TKYGSKVYII HRYDAFNASK IMQORALSNP KIDVINSSSV 250
RYN-I-S-Tagged Protein DSAMEEANFL TKYGSKVYII HRYDAFNASK IMQORALSNP KIDVINSSSV
RYN-L-S-Tagged Protein DSAMEEANFL TKYGSKVYII HRYDAFNASK IMQORALSNP KIDVINSSSV
RYN-M-S-Tagged Protein DSAMEEANFL TKYGSKVYII HRYDAFNASK IMQORALSNP KIDVINSSSV
251
RRR-WT-S-Tagged Protein VEAYGDGERD VLGLGKVNK VTDGVSCLKV SGLPFAIGHE PATKFLDGGV 300
RYN-I-S-Tagged Protein VEAYGDGERD VLGLGKVNK VTDGVSCLKV SGLPFAIGHE PATKFLDGGV
RYN-L-S-Tagged Protein VEAYGDGERD VLGLGKVNK VTDGVSCLKV SGLPFAIGHE PATKFLDGGV
RYN-M-S-Tagged Protein VEAYGDGERD VLGLGKVNK VTDGVSCLKV SGLPFAIGHE PATKFLDGGV
301
RRR-WT-S-Tagged Protein ELDSGDYVVT KPGTTQTSVP GVFAAGDVQD KKYRQAITAA GTGCMALDA 350
RYN-I-S-Tagged Protein ELDSGDYVVT KPGTTQTSVP GVFAAGDVQD KKYRQAITAA GTGCMALDA
RYN-L-S-Tagged Protein ELDSGDYVVT KPGTTQTSVP GVFAAGDVQD KKYRQAITAA GTGCMALDA
RYN-M-S-Tagged Protein ELDSGDYVVT KPGTTQTSVP GVFAAGDVQD KKYRQAITAA GTGCMALDA
351
RRR-WT-S-Tagged Protein EHYLQBIQSG QGKSD 365
RYN-I-S-Tagged Protein EHYLQBIQSG QGKSD
RYN-L-S-Tagged Protein EHYLQBIQSG QGKSD
RYN-M-S-Tagged Protein EHYLQBIQSG QGKSD

```

FIGURE 17B

【図 17 A】

使用した発現ベクターによってコードされるような、NTR WTおよび新変異体タンパク質配列のアラインメント。アミノ酸位置の記載のために以下の色凡例を参照されたい。

```

RRR-WT: MKETAAAKE RQMDSDLG TLVPRGSMAD IRNGLESTN TRLCIVGSSP 100
RYN-I: MKETAAAKE RQMDSDLG TLVPRGSMAD IRNGLESTN TRLCIVGSSP 100
RYN-L: MKETAAAKE RQMDSDLG TLVPRGSMAD IRNGLESTN TRLCIVGSSP 100
RYN-M: MKETAAAKE RQMDSDLG TLVPRGSMAD IRNGLESTN TRLCIVGSSP 100
RRR-WT: AAHTAAIYAA RAELEPLIFE GWANDIAPG GOLTTTIDE NFGPFPEGIL 100
RYN-I: AAHTAAIYAA RAELEPLIFE GWANDIAPG GOLTTTIDE NFGPFPEGIL 100
RYN-L: AAHTAAIYAA RAELEPLIFE GWANDIAPG GOLTTTIDE NFGPFPEGIL 100
RYN-M: AAHTAAIYAA RAELEPLIFE GWANDIAPG GOLTTTIDE NFGPFPEGIL 100
RRR-WT: GVELTDKFRK QSERPTTIF TETVTKVDFS SKPKLFIDS KAILADAVIL 100
RYN-I: GVELTDKFRK QSERPTTIF TETVTKVDFS SKPKLFIDS KAILADAVIL 100
RYN-L: GVELTDKFRK QSERPTTIF TETVTKVDFS SKPKLFIDS KAILADAVIL 100
RYN-M: GVELTDKFRK QSERPTTIF TETVTKVDFS SKPKLFIDS KAILADAVIL 100
RRR-WT: ATGAVAKRLS FVGSBGSOG FWRNGISACA VCDGAPFIR NKPLAVIGGG 100
RYN-I: ATGAVAKRLS FVGSBGSOG FWRNGISACA VCDGAPFIR NKPLAVIGGG 100
RYN-L: ATGAVAKRLS FVGSBGSOG FWRNGISACA VCDGAPFIR NKPLAVIGGG 100
RYN-M: ATGAVAKRLS FVGSBGSOG FWRNGISACA VCDGAPFIR NKPLAVIGGG 100
RRR-WT: DSAMEEANFL TKYGSKVYII HRYDAFNASK IMQORALSNP KIDVINSSSV 100
RYN-I: DSAMEEANFL TKYGSKVYII HRYDAFNASK IMQORALSNP KIDVINSSSV 100
RYN-L: DSAMEEANFL TKYGSKVYII HRYDAFNASK IMQORALSNP KIDVINSSSV 100
RYN-M: DSAMEEANFL TKYGSKVYII HRYDAFNASK IMQORALSNP KIDVINSSSV 100
RRR-WT: VEAYGDGERD VLGLGKVNK VTDGVSCLKV SGLPFAIGHE PATKFLDGGV 100
RYN-I: VEAYGDGERD VLGLGKVNK VTDGVSCLKV SGLPFAIGHE PATKFLDGGV 100
RYN-L: VEAYGDGERD VLGLGKVNK VTDGVSCLKV SGLPFAIGHE PATKFLDGGV 100
RYN-M: VEAYGDGERD VLGLGKVNK VTDGVSCLKV SGLPFAIGHE PATKFLDGGV 100
RRR-WT: ELDSGDYVVT KPGTTQTSVP GVFAAGDVQD KKYRQAITAA GTGCMALDA 100
RYN-I: ELDSGDYVVT KPGTTQTSVP GVFAAGDVQD KKYRQAITAA GTGCMALDA 100
RYN-L: ELDSGDYVVT KPGTTQTSVP GVFAAGDVQD KKYRQAITAA GTGCMALDA 100
RYN-M: ELDSGDYVVT KPGTTQTSVP GVFAAGDVQD KKYRQAITAA GTGCMALDA 100
RRR-WT: EHYLQBIQSG QGKSD 100
RYN-I: EHYLQBIQSG QGKSD 100
RYN-L: EHYLQBIQSG QGKSD 100
RYN-M: EHYLQBIQSG QGKSD 100

```

凡例

T R-2 ライブラリで設計された非RRR位置

RRR-WT-S-Tagged Protein

FIGURE 17A

【図 18】

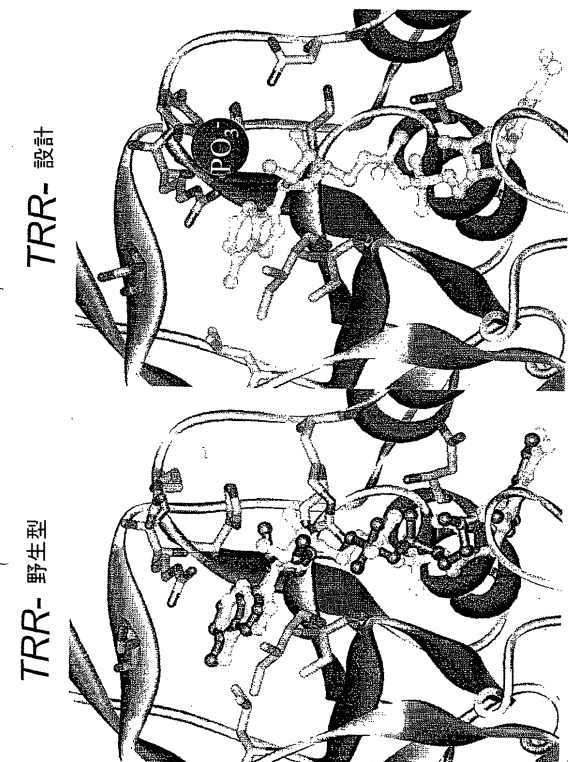


FIGURE 18

【図 2 1 - 1】

Figure 21A

>sp|P09625|TRXB_ECOLI トレッド'キリンダ'クラーゼ (EC 1.6.4.5) (TRXR) - Escherichia coli, 及び Escherichia coli O157:H7
 GTTKHKKLLILGSGPAGYTAYVAAARAKLQPVLTIGMEKGGQLTTTTEVENWPGDNDLT
 GPLLMERMEHATKFEZEIFDHNKVDLQNRPRFNDGNGEYTCDALIATGASARYLG
 LPSEAFKRGVACATCDGFFYRNQKVAVIGGNTAVBEALYLGNISAEVHLHRRDGF
 RAKLLIKRLNDKVENGNILHTNRTLESEVTDGQMGVTEGRLDRTQNSDNIESLDVAGLF
 VALGHSPTAIFPGQLLENGYIKVQSGIHGNATQTSIFGVPFAGDVMDHIYRQAITSG
 TGCMAALDAERYLQGLADAK

Figure 21B

>sp|P08080|TRXB_BACSU トレッド'キリンダ'クラーゼ (EC 1.6.4.5) (TRXR) (一般的なストレスタンバク菌 35) (GSF35) - Bacillus subtilis.
 SEETKVDVILIGSGPAGMTAVYTSRANSLTALIERGIFPGGQMANTEOVENYPGFESILG
 PELSNKMFERRAKFGAEVAYGDIKEVIDCKYKVKAGSEYKARAVILAGAEYKIKGV
 PGEKELGGGRGSYCAVCDGAFFKGKELVVVGSGDSAVEEGVYLTFASKVTIVHRRDKLR
 AGSILQARAFDNKVDFLANKNTEKLEHEBNGKNGVTLVDVTGEESEFKTGVFYIYGM
 LFLSKPFENLIGITNEBGYIETNDRMTKVEGIFPAGDIRKSLRQIVTATGDSIAAQSV
 QHYVEELQETLAKLR

Figure 21C

>sp|P46843|TRXB_MYCLE 二機能性 トレッド'キリンダ'クラーゼ / トレッド'キノ [Includes: トレッド'キリンダ'クラーゼ (EC 1.6.4.5) (TRXR); トレッド'キノ] - Mycobacterium leprae.
 MNTPPSAHETIHEVIVIGSGPAGYPAALYAARQLPLVPEGTSFGGALMTTTEVENYPG
 FRNKTGCELMNDMDALPAGALKTEDOVESLACKEKSVYTAGQTYQARAVILAMG
 TSVRYLQIPGQDELLGRGVACATCDGSPFRNQDIAVIGGDSAMBEALTLFEPARSVL
 VHRDEFRASKIMLGRANNDIKIFITNHTVAVNGYITVTGRLRLNTTTEGETTLVVTG
 VFVALGHEPRSSLYSDVDVIDPDGYLVKRGTTSTSDMDGVFAAGLDVDRTYRQAITAAGS
 GCAALDAERWLAHAGSEKANTTETPGDVDSITDTTWTAMTDARNGVTIEVTDASFF
 ADVLSNKFVLDVDFRWCSPCKVAVFLBEIASGQRNLTVAKLDDVTNFEMAREFQVV
 SIPTMILFQGGQPVKRIYGAAGKAAALLRLSLDVPFHL

Figure 21D

>sp|P51978|TRXB_NEUCR トレッド'キリンダ'クラーゼ (EC 1.6.4.5) - Neurospora crassa.
 MHSKVVLIGSGPAAHTAAIYLARAEKPLVYBGFMANIAAGQLTTTTEIENFPFGPDG
 INGGEIMDRMKAQSPFGTQIISFTVAKVULSARPKYATWEPFEHTADSILATGAS
 ARRLHLPGKEEYNGIGISACAVCDGAVPFRNKHVLVIGGDSARAEKMLPKYGSHTV
 LVKRDHLRASSINARILNHEKVTIVRTNVTVALEAKDGGKLMHLLVVDVTTGKEETLEA
 NGLFYALGHDPATALVKQLETDADGVVTKPGTTLTSVEGVFAAGDVQDKRYRQAITSA
 GTGCMALDAEKFLSEHEETFAEHRDTSVQNG

Figure 21E

>sp|P29509|TRB1_YEAST トレッド'キリンダ'クラーゼ 1 (EC 1.6.4.5) - Saccharomyces cerevisiae (Bakerの酵母).
 VHNKVTLIGSGPAAHTAAIYLARAEIKPLVYEGMMANGIAAGQLTTTTEIENFPFGPDG
 IAGSELMDRMKQSTKFGTEIITETVSKVDLSKPKPLWTEFNEDEPVTDAIILATGA
 SAKRMHLPGESTVWKGISACAVCDGAVIFRNKPLAVIGGDSACBAQFLTKYGSKVF
 MLVRKDHRASTIMCKRAEKNKEILELYNTVVALEAKDGGKLMHLLVVDVTTGKEETLEA
 SGLFYAIGHTPATKIYAGQVDTDEAGYIKTVFPGSSITSVFPGFAGDVQDSKYRQAITSA
 GSGCMALDAEKYLTSL

Figure 21F

>sp|P38816|TRB2_YEAST トレッド'キリンダ'クラーゼ 2, ミトコンドリア前駆体 (EC 1.6.4.5) - Saccharomyces cerevisiae (Bakerの酵母).
 MIKHIVSPFRTNFVIGSKSVLSRMHKKVTIGSGPAAHTAAIYLARAEKPLTYEGMMA
 NGIAGGQLTTTTEIENFPFGFESILSGSBIEMRMKQSAKFGTNIIITETVSKVDLSKPKP
 RLWTEFNEDEPVTDAIILATGASAKRMHLPGESTVWQGISACAVCDGAVIFRNKPL
 AVIGGDSACBAEFLTKYASKVYLVKRDHPRASVIMGRKRWIKRWIIVLENTVALEAK
 GDGKLMHLLRINKTKSNVENDLLEVNLFLYALGHSFATDITVKGQVDEETGYIKTVPGSSL
 TSVFPGFAGDVQDSRYRQAVTSAGSGCIILDAERYLSAQE

Figure 21G

>sp|Q39243|TRB1_ARATH トレッド'キリンダ'クラーゼ 1 (EC 1.6.4.5) (NADPH-依存性 トレッド'キリンダ'クラーゼ 1) (NTR 1) - Arabidopsis thaliana (マウス耳クレス).

【図 2 1 - 2】

MNGLETHNTRLCIVGSGPAAHTAAIYARAEKPLFEGWMANDIAPGGQLTTTDDVENF
 NSPFEGILQVELVETKFKQDSREGTITETVTKVDFPSKPKPLFTDSKALLADAVILAT
 GAVAKRLSFYVSGSEASGGFWNRGTSACAVCDGAAPIFRNKPLAVIGGDSAMBEALFLK
 YGSKVYIHRDDAFRAKIMQQRALSNPKIDIVNWSVVEAYGGERDVLGGLKVENVVT
 GDVSDLKVSGLFPAIGHPEPATKFLDGGVELSDSGYVVTGPTQTSTVFGVFAAGDVQDK
 YRQAITAAGTGCMAALDAERYLQEGISQGGKSD

Figure 21H

>sp|Q39242|TRB2_ARATH トレッド'キリンダ'クラーゼ 2 (EC 1.6.4.5) (NADPH-依存性 トレッド'キリンダ'クラーゼ 2) (NTR 2) - Arabidopsis thaliana (マウス耳クレス).
 MCWISMSQSRFLIKSLFSTAGGFLGSLSNPFSLATAFSSSSSSSSAAAAVDMETHKTK
 VCIIVGSGPAAHTAAIYASAEKPLFLLFEGWMANDIAPGGQLTTTDDVENFPFGFEGILGI
 DIVKFRKQSERPQTITETVTKVDFPSKPKPLDTSRVLADSVLSTGAVAKRLSFT
 GSGEGNGGFWNRGTSACAVCDGAAPIFRNKPLVIGGDSAMBEALFLKYSKYIHR
 RDTFRASKIMQQRALSNPKIENVNWSAVVEAYGDERNGRVLGGLKVINVTGDSDLKVS
 LFFAIGHPEPATFLDQGLEDEDGYVVTGPTGTTSTVVGFAAGDVQDKRYRQAITAAGT
 GCMALDAERYLQEGISQGGKSD

Figure 21I

>sp|Q16801|TRXB_HUMAN トレッド'キリンダ'クラーゼ (EC 1.6.4.5) - Homo sapiens (ヒト).
 MNGPEDLPKSYDYDLIIIGSGSGLAARAKAAQYKGVMLVDFVTPPLGTRWGLSGTCV
 NVGCIPIKILMBAQALIGQALQDSRNYGWKVEETVKHDMWDMLEAVQNHGISLNGVRYAL
 REKVVYENATGQFIPHRIKATNNKGEKITYSABSFLIATGERPRYLGI PGDKYCISS
 DLFSPFLPCPKTLVVGASVVALCBAGFLASIGLVTWVRSILLAGFDQDAKILGBN
 EBHGKIFIRQFPVPIKVEQIEAGTPGRILRVVAQSTNSEEIIIEGEYNTVMAIGRDACTRI
 GLETVGVINERKTGKIPVTEDEQNTVPYIYAGIDILEKVELTPVAIQGRLLAQRLYAG
 STVKCOYENVPTVTFPLEYACGLSEKAVEKTEENIEVYHSYFPLEWPTISPDNNK
 CIAKLICHTKDRERVVGFVLGPNAGSEVTGGFAALKQGLTKRQLDSTIGIHPVCARVFT
 TLSVTKRSGASTLQAGC

【図 2 1 - 3】

Figure 21J

trxB 由来 Methanococcus jannaschii (gil592167):

MIHDTIIGAGPGLTAGIYAMRGKLNALCIEKENAGGRIAEAGIVENYPGFEIRGYELAKFP
 KNHAEPKPLPIYDVEIKIETKRPFKVITFNSEYLTAKTIVIAQTGPKPLGLNEDKFIGRGIS
 YCTMCDAPFYLNKEVIVIGRDTFAIMSAINLKDIAKKVIVITDKSELKAAESIMLDLKEANNV
 EIIYNAPLEIVGEERABGVKIVNGKEEIIKADGIFISLGHVPNTEFLKDSIGLELKKGFIKT
 DENCRNTNIDGIYAVGDVVRGGVMQVAKYAGDGVAMANIIKYLQKL

Figure 21K

trxB 由来 Amhaeoglobusfulgidus (gil2649006):

MYDVAILIGSGPAGLTAALYSARYGLKTVFFSTVDVPSQLSAAKIENYPGFEGSGMELLERKME
 QAVKAGAEWKLKVERVERNRGTFTVAEGGEYAKALIVATGGKHKEAGIBGESAFIGRGVSY
 CATCDGNPFRRKGVIVYSGSKAIEDAIIYLDIGCEVTIVSRTPSPRAEKALVEEVEKRGIFVH
 YSTTRIKLIGSGKVEKYVAVNRKKKEEFLADGIFVIAIGMRPATDVVAELGVSRDSMGYIKVD
 KQRTNRVGVFAAGDCCDNPLKQVVTAQCDGAVAAYSATKYLTS.

【国際公開パンフレット】

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau(43) International Publication Date
14 November 2002 (14.11.2002)

PCT

(10) International Publication Number
WO 02/090300 A2

- (51) International Patent Classification: C07C Torrey Mesa Research Institute/Syngenta, 3115 Merryfield Row, San Diego, CA 92121-1125 (US); MUCHHAL, Umesh; c/o Torrey Mesa Research Institute/Syngenta, 3115 Merryfield Row, San Diego, CA 92121-1125 (US).
- (21) International Application Number: PCT/US02/14358
- (22) International Filing Date: 6 May 2002 (06.05.2002)
- (25) Filing Language: English
- (26) Publication Language: English
- (30) Priority Data:
4 May 2001 (04.05.2001) US 60/289,029
5 April 2002 (05.04.2002) US 60/370,609
N/A 29 April 2002 (29.04.2002) US
- (74) Agents: TRECARDIN, Richard, F. et al.; Dorsey & Whitney LLP, Suite 3400, 4 Embarcadero Center, San Francisco, CA 94111-4187 (US).
- (81) Designated States (national): AL, AG, AI, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GR, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, UZ, VN, YU, ZA, ZM, ZW.
- (84) Designated States (regional): ARIPO patent (GI, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).
- Published:
— without international search report and to be republished upon receipt of that report
- For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.
- (71) Applicants: XENCOR [US/US]; 111 West Lemon Avenue, Menlo Park, CA 94016 (US); SYNGENTA PARTICIPATIONS AG [CH/CH]; Schwarzwaldallee 215, CH-4058 Basel (CH).
- (72) Inventors: BRIGGS, Steven, P.; c/o Torrey Mesa Research Institute/Syngenta, 3115 Merryfield Row, San Diego, CA 92121-1125 (US); DALMIA, Bipin, K.; c/o Torrey Mesa Research Institute/Syngenta, 3115 Merryfield Row, San Diego, CA 92121-1125 (US); DEL VAL, Gregory; c/o Torrey Mesa Research Institute/Syngenta, 3115 Merryfield Row, San Diego, CA 92121-1125 (US); DESJARLAIS, John, R.; c/o Torrey Mesa Research Institute/Syngenta, 3115 Merryfield Row, San Diego, CA 92121-1125 (US); HEIFETZ, Peter; c/o Torrey Mesa Research Institute/Syngenta, 3115 Merryfield Row, San Diego, CA 92121-1125 (US); LUGINBUHL, Peter; c/o



WO 02/090300 A2

(54) Title: NUCLEIC ACIDS AND PROTEINS WITH THIOREDOXIN REDUCTASE ACTIVITY

(57) Abstract: The present invention relates to the use of a variety of methods for generating functional thioredoxin reductase variants in which at least one physical, chemical or biological property of the variant is altered in a specific and desired manner when compared to the wild-type protein.

WO 02/090300

PCT/US02/14358

5 NUCLEIC ACIDS AND PROTEINS WITH THIOREDOXIN REDUCTASE ACTIVITY

This application claims the benefit of the filing date of U.S.S.N. 60/289,029, filed May 4, 2001, U.S.S.N. 60/370,609, filed April 5, 2002, and the provisional application by Desjarlais and Muchhal, entitled "Novel Nucleic Acids and Proteins with Thioredoxin Reductase Activity", filed April 29, 2002,
10 serial number not assigned.

FIELD OF THE INVENTION

The present invention relates to the use of a variety of methods for generating functional thioredoxin reductase variants in which at least one physical, chemical or biological property of the variant is altered in a specific and desired manner when compared to the wild-type protein.

BACKGROUND OF THE INVENTION

20 Thioredoxin, a small dithiol protein, is a specific reductant for major food proteins, allergenic proteins and particularly allergenic proteins present in widely used foods from animal and plant sources. Most proteins having disulfide (S-S) bonds are reduced to the sulfhydryl (SH) level by thioredoxin. These proteins are allergenically active and less digestible in the oxidized (S-S) state. When reduced (SH state), they lose their allergenicity and/or become more digestible. Of importance is the thioredoxin
25 reduction of disulfide bonds in proteins such as albumins, globulins, gliadins, thionins, and the glutenins found in many seeds and cereals, and also a number of proteins found in milk. See, for example, Kiss, F. et al. (1991), *Arch. Biochem. Biophys.* 287:337-340; Johnson, T. C. et al. (1987), *Plant Physiol.* 85:446-451; Kasarda, D. D. et al. (1976), *Adv. Cer. Sci. Tech.* 1:158-236; and Osborne, T. B. et al. (1993), *Amer. Chem. J.* 15:392-471; Shewry, P. R. et al. (1985), *Adv. Cer. Sci. Tech.* 7:1-
30 83; Dahle, L. K. et al. (1966), *Cereal Chem.* 43:682-688; Garcia-Olmedo, F. et al. (1987), *Oxford Surveys of Plant Molecular and Cell Biology* 4:275-335; Birk, Y. (1976), *Meth. Enzymol.* 45:695-739, and Laskowski, M., Jr. et al. (1980), *Ann. Rev. Biochem.* 49:593-626; Weseleake, R. J. et al. (1983), *Plant Physiol.* 72:809-812; Birk, Y. (1985), *Int. J. Peptide Protein Res.* 25:113-131, and Birk, Y. (1976), *Meth. Enzymol.* 45:695-739; Birk, Y. (1985), *Int. J. Peptide Protein Res.* 25:113-131.

35 In addition, thioredoxin reduces the disulfide bonds in many toxic proteins, such as those found in snakes (Yang, C. C. (1967) *Biochim. Biophys. Acta.* 133:346-355; Howard, B. D. et al. (1977) *Biochemistry* 16:122-125), bees, scorpions (Watt, D. D. et al. (1972) *Toxicon* 10:173-181), the bacterial neurotoxins tetanus and botulinum (Schiavo, G. et al. (1990) *Infection and Immunity*

WO 02/090300

PCT/US02/14358

58:4136-4141; Kistner, A. et al. (1992) *Naunyn-Schmiedeborg's Arch Pharmacol* 345:227-234), and thereby reduces or in some instances eliminates their toxicity altogether.

Thioredoxin achieves this reduction when activated (reduced) either by nicotinamide adenine dinucleotide phosphate (NADPH) via NADP-thioredoxin reductase (physiological conditions) or by dithiothreitol, a chemical reductant. See, for example, U.S. Patent No. 5,952,034, incorporated herein by references in its entirety. Skin tests and feeding experiments carried out with sensitized dogs have shown that treatment of the food with reduced thioredoxin prior to ingestion eliminates or decreases the allergenicity of the food. Studies have also shown increased digestion of food and food proteins by pepsin and trypsin following reduction by thioredoxin.

Thus, it would be desirable to develop an efficient, low cost method of using thioredoxin reductase to reduce the toxicity of toxic proteins, reduce the allergenicity of food, and increase the digestibility of food.

SUMMARY OF THE INVENTION

In accordance with the objects outlined above, the present invention provides a method for altering the cofactor specificity of thioredoxin reductase comprising inputting a set of coordinates for a thioredoxin reductase scaffold protein comprising amino acid positions; applying at least one protein design cycle, and generating a set of candidate variant proteins with altered cofactor dependency. Preferably, the scaffold protein is selected from the group of organisms consisting of *E. coli*, *Bacillus subtilis*, *Mycobacterium leprae*, *Sarccharomyces*, *Neurospora crassa*, *Arabidopsis*, and human.

In an additional aspect, the cofactor specificity of the variant TR protein is NADPH or NADH. Preferably, the cofactor specificity is switched to NADH. In addition, other TR variants are generated that preferentially bind NADPH compared to NADH, preferentially bind NADN compared to NADPH, bind both cofactors equally. In other embodiments, the catalytic efficiency for one or the other cofactors or both is altered.

In an additional aspect the variant TR proteins have amino acid substitutions selected from the group of substitutions consisting of RA4W, RA5L, R A5M, R A5I, R A5F, R A5V, R A5Y, RA5A, RA5S, RA5C, RA5T, RA6T, R A6S, R A6Q, R A6G, and R A6N, RA6D, RA6M, and RA6E.

In an additional aspect, the present invention provides a method for altering the substrate specificity of TR protein comprising inputting a set of coordinates for a thioredoxin reductase scaffold protein comprising amino acid positions; applying at least one protein design cycle, and generating a set of candidate variant proteins with altered substrate specificity.

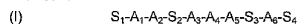
WO 02/090300

PCT/US02/14358

In an additional aspect, the present invention provides a method for altering the cofactor specificity of a target protein comprising inputting a set of coordinates for a thioredoxin reductase scaffold protein comprising amino acid positions; applying at least one protein design cycle, and generating a set of candidate variant proteins with altered cofactor specificity.

5

In an additional embodiment, the present invention provides a variant thioredoxin reductase (TR) protein comprising an isolated polypeptide molecule of Formula I



wherein

- 10 a) S_1 comprises a polypeptide sequence selected from the group consisting of SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, and SEQ ID NO:7, or a sequence having substantial similarity thereto;
- b) S_2 comprises a polypeptide sequence selected from the group consisting of SEQ ID NO:8, SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, and SEQ ID NO:14, or a sequence having substantial similarity thereto;
- 15 c) S_3 comprises a polypeptide sequence selected from the group consisting of SEQ ID NO:15, SEQ ID NO:16, SEQ ID NO:17, SEQ ID NO:18, SEQ ID NO:19, SEQ ID NO:20, and SEQ ID NO:21, or a sequence having substantial similarity thereto;
- d) S_4 comprises a polypeptide sequence selected from the group consisting of SEQ ID NO:22, SEQ ID NO:23, SEQ ID NO:24, SEQ ID NO:25, SEQ ID NO:26, SEQ ID NO:27, and SEQ ID NO:28, or a sequence having substantial similarity thereto;
- 20 e) A_1 is an amino acid moiety selected from the group consisting of serine, valine, glycine, alanine, leucine, isoleucine, methionine, phenylalanine, and tryptophan;
- f) A_2 is an amino acid moiety selected from the group consisting of alanine, glycine, valine, leucine, isoleucine, methionine, phenylalanine, and tryptophan;
- 25 g) A_3 is an amino acid moiety selected from the group consisting of histidine, aspartic acid, glutamic acid, arginine, leucine, serine, threonine, cysteine, asparagine, glutamine, and tyrosine;
- h) A_4 is an amino acid moiety selected from the group consisting of arginine, alanine, glycine, valine, leucine, isoleucine, methionine, phenylalanine, and tryptophan;
- 30 i) A_5 is an amino acid moiety selected from the group consisting of arginine, asparagine, glutamine, aspartic acid, glutamic acid, cysteine, serine, threonine, and lysine;
- j) A_6 is an amino acid moiety selected from the group consisting of arginine, glutamic acid, asparagine, glutamine, aspartic acid, cysteine, serine, threonine, and lysine;
- 35

provided that at least

- A_1 is not serine;
 A_2 is not alanine;
 A_3 is not histidine;

WO 02/090300

PCT/US02/14358

A₄ is not arginine;
 A₅ is not arginine; or
 A₆ is not arginine.

- 5 In an additional aspect, the present invention provides a method for altering the oil content of plant cells comprising introducing an expression cassette comprising a promoter functional in a plant cell operably linked to a DNA molecule encoding a modified thioredoxin reductase (TR) protein according to claim 1 or 22 comprising an amino terminal chloroplast transit peptide, into the cells of a plant so as to yield transformed plant cells; and regenerating said transformed plant cells to provide a
- 10 differentiated transformed plant, wherein expression of the DNA molecule encoding the modified TR protein in said plant alters the co-factor specificity compared to the untransformed plant.

BRIEF DESCRIPTION OF THE DRAWINGS

- 15 Figure 1 depicts the reaction catalyzed by thioredoxin reductases.

Figure 2 depicts the active site pocket of reductases from a number of species is highly conserved. Figure 2A lists some of the most common TR sequences. The first column lists the Genbank ID number, A1 through A6 refers to the amino acids defined in Formula I (described below), S2 and S3 are sequence domains separating A1 through A6 and are also defined in Formula I.

- 20

Figure 2B lists some of the common glutathione reductase sequences.

- Figures 2C and 2D represent the natural sequence diversity at each of the defined positions grouped according to organism.
- 25

Figure 2E lists known cofactor specificity and known amino acid placement.

- Figure 3 depicts various sequences that may be used in Formula I.
- 30

Figure 4 provides an overview of the high throughput TR screening methods.

Figure 5 depicts protein purification strategies.

- 35 Figure 6 depicts the kinetics of Arabidopsis NTR wild-type reductase with NAD(P)H.

Figure 7 depicts variants obtained from the NTR-1 Library 1.

Figure 8 depicts variants obtained from the NTR-1 Library 2.

WO 02/090300

PCT/US02/14358

Figures 9A and 9B depict the designed positions and the docked co-factor from NTR-1 Library 1 and NTR-1 Library 2.

- 5 Figure 10 depicts the summary of results from the screening of variants from 4 computational libraries.

Figures 11A and B depict the kinetic parameters for 2 variants versus wild-type TR.

- 10 Figure 12 depicts a summary of the best variants obtained from the NTR-1 library 2 design.

Figures 13A and B summarize the activity of variants obtained from a high complexity random RRR library. A summary of the variants obtained from this library is found in Figure 13C.

- 15 Figure 14 depicts a computational model for two of the clones.

Figure 15 summarizes the enzymatic activities and kinetic parameters for some of the variants.

Figure 16A depicts the nucleic acid sequence for the WVR variant.

20

Figure 16B depicts the nucleic acid sequence for the WMG variant.

Figure 16C depicts the nucleic acid sequence for the WIS variant.

- 25 Figure 16D depicts the nucleic acid sequence for the WMS variant.

Figure 16E depicts the nucleic acid sequence for the WLS variant.

Figure 16F depicts the nucleic acid sequence for the WRT variant.

30

Figure 16G depicts the nucleic acid sequence for the RYN variant.

Figure 16H depicts the nucleic acid sequence for the RYN-A variant.

- 35 Figure 16I depicts the nucleic acid sequence for the RFN variant.

Figure 16J depicts the RRR-WT nucleic acid sequence.

Figure 16K depicts the nucleic acid sequence for the VWG variant.

WO 02/090300

PCT/US02/14358

Figure 16L depicts the nucleic acid sequence for the WRS variant.

Figure 16M depicts the nucleic acid sequence for the WFQ variant.

5 Figure 16N depicts the nucleic acid sequence for the NTR wild-type protein.

Figure 16O depicts the nucleic acid sequence for the RYN-M variant.

Figure 16P depicts the nucleic acid sequence for the RYN-L variant.

10

Figure 16Q depicts the nucleic acid sequence for the RYN-I variant.

Figures 17A and B depict the alignment of the Arabidopsis NTR wild-type protein with several of the variants.

15

Figure 18 is a computational representation of the critical RRR to RYN change described in Example 1.

20 Figure 19 depicts a small sample of NAD conformations culled from the protein databank. The ball-and-stick model is the NAD_TDF conformer, which has a different ribose pucker than most of the others.

Figure 20 depicts the library positions utilized in PDA simulations and generation of libraries 1 and 2.

25 Figure 21 depicts the sequence alignment of several wild-type TR proteins. Sequences correspond to the following: 1) |P09625|TRXB_ECOLI; 2) |P80880|TRXB_BACSU; 3) |P46843|TRXB_MYCLE; 4) |P51978|TRXB_NEUCR; 5) |P29509|TRB1_YEAST; 6) |P38816|TRB2_YEAST; 7) |Q39243|TRB1_ARATH; 8) |Q39242|TRB2_ARATH; and, 9) |Q16881|TRXB_HUMAN.

30 Figure 22 depicts the amino acid sequences of several wild-type TR proteins. Sequences correspond to the following: A) |P09625|TRXB_ECOLI; B) |P80880|TRXB_BACSU; C) |P46843|TRXB_MYCLE; D) |P51978|TRXB_NEUCR; E) |P29509|TRB1_YEAST; F) |P38816|TRB2_YEAST; G) |Q39243|TRB1_ARATH; H) |Q39242|TRB2_ARATH; and, I) |Q16881|TRXB_HUMAN.

35

DETAILED DESCRIPTION OF THE INVENTION

The present invention is directed to the generation of variant proteins and nucleic acids that exhibit altered cofactor specificity. The variant proteins may be generated using a number of different approaches, such as conventional mutagenesis approaches and computational processing

WO 02/090300

PCT/US02/14358

approaches. Computational processing approaches have been previously described in U.S. Patent Nos. 6,188,965 and 6,296,312, U.S.S.N.s 09/419,351, 09/782,004, 09/927,79, and 09/877,695; all of which are expressly incorporated herein by reference in their entirety. In general, these applications describe a variety of computational modeling systems that allow the generation of extremely stable proteins. In this way, variants of wild-type proteins are generated that exhibit altered cofactor specificity as compared to wild-type proteins.

The methods of the present invention can be applied to any enzyme that exhibits a preference for one cofactor over another. For example, enzyme reductases often exhibit a preference for one cofactor versus another. In addition, the methods of the present invention can be applied to change the substrate specificity of a target protein.

In particular, the methods of the present invention can be used to change the cofactor preference from NADPH to NADH. NADPH is an expensive reductant. Its expense has prohibited the wide use of thioredoxin systems in reducing food allergens and venom treatments. Thus, there is a need in the art to find other systems that achieve the same results as the use of NADP-thioredoxin reductase reductants but at lower costs. One such system, would be to generate variants of thioredoxin reductase with altered cofactor specificity.

According to the present invention provides methods for altering the cofactor specificity of a target protein. By "altering" herein or grammatical equivalents thereof in the context of a polypeptide, as used herein, further refers to any characteristic or attribute of a polypeptide that can be selected or detected and compared to the corresponding property of a naturally occurring protein. These properties include, but are not limited to cofactor specificity, cytotoxic activity; oxidative stability, substrate specificity, substrate binding or catalytic activity, thermal stability, alkaline stability, pH activity profile, resistance to proteolytic degradation, kinetic association (K_{on}) and dissociation (K_{off}) rate, protein folding, inducing an immune response, ability to bind to a ligand, ability to bind to a receptor, ability to be secreted, ability to be displayed on the surface of a cell, ability to oligomerize, ability to signal, ability to stimulate cell proliferation, ability to inhibit cell proliferation, ability to induce apoptosis, ability to be modified by phosphorylation or glycosylation, ability to treat disease.

Unless otherwise specified, a substantial change in any of the above-listed properties, when comparing the property of a variant polypeptide of the present invention to the property of a target protein or wild-type protein is preferably at least a 20%, more preferably, 50%, more preferably at least a 2-fold increase or decrease.

By "cofactor specificity" herein is meant changing the cofactor preference of an enzyme. By "cofactor" herein is meant coenzymes, such as NADPH, NADH, that participate in oxidation/reduction reactions. Thus, if a target protein exhibits a preference for one cofactor over another, the methods of the

WO 02/090300

PCT/US02/14358

present invention may be used to alter the cofactor preference of the target enzyme, such that the preference for the less favored cofactor is increased by 20%, 50%, 100%, 300%, 500%, 1000%, up to 2000%. For example, a number of reductase enzymes favor NADPH over NADH (see WO 02/22526; WO 02.29019; Mittl, PR., et al., (1994) Protein Sci., 3: 1504-14; Bantia, S., et al., (2002) Protein Eng., 15:131-140; all of which are hereby incorporated by reference in their entirety). As the availability of NADPH is often limiting, both *in vivo* and *in vitro*, the overall activity of target protein is often limited. For target proteins that prefer NADPH as a cofactor, it would be desirable to alter the cofactor specificity of the target protein to a cofactor that is more readily available, such as NADH.

- 10 In a preferred embodiment, the cofactor specificity of the target protein is switched. By "switched" herein is meant, that the cofactor preference (e.g. affinity) of a target protein is changed to another cofactor. Preferably, in one embodiment, by switching cofactor specificity, activity with the cofactor preferred by the wild-type enzyme is reduced, while the activity with the less preferred cofactor is increased. For example, if a target protein prefers NADPH, switching the preference to NADH would result in the variant TR having at least 50% of native NADPH dependent activity using NADH. More preferably, the variant TRs will have at least 75% of native NADPH dependent activity using NADH. More preferably the variant TRs will have 85%, 95%, up to 100% of native NADPH activity using NADH. Alternatively, in another embodiment, the alternate cofactor affinity is increased without a decrease in preferred cofactor affinity. In yet other embodiments, the cofactor affinity for both factors is changed simultaneously.

- 25 In a preferred, the catalytic efficiency of the target protein for a cofactor is enhanced. By "catalytic efficiency" herein is meant the activity with the cofactor is significantly improved. Catalytic efficiency may be improved for either the preferred cofactor or, in those embodiments where the cofactor specificity is altered the catalytic efficiency with the altered cofactor may be improved.

- 30 In a preferred embodiment, the binding affinity of the target protein for a cofactor is enhanced. A change in binding affinity is evidenced by at least a 5% or greater increase or decrease in binding affinity compared to the wild-type target protein. In certain embodiments, variant proteins of the present invention may show greater than 100 times more affinity for one cofactor than for another, while in other embodiments the variant protein may show greater than 50 times more affinity for one cofactor than for another, or the variant protein may show greater than 25 times more affinity for one cofactor than another.

- 35 In a preferred embodiment, the substrate specificity of the target protein is altered. For example, if a target protein typically acts on a substrate from the same species, the substrate specificity of the target protein may be changed such that the variant protein acts on substrates from other species.

WO 02/090300

PCT/US02/14358

Accordingly, the present invention is directed to methods for altering the cofactor specificity of target protein. By "target protein" or "scaffold protein" or grammatical equivalents herein is meant at least two covalently attached amino acids, which includes proteins, polypeptides, oligopeptides and peptides. The protein may be made up of naturally occurring amino acids and peptide bonds, or synthetic peptidomimetic structures, i.e., "analogs" such as peptoids [see Simon et al., Proc. Natl. Acad. Sci. U.S.A. 89(20):9367-71 (1992)], generally depending on the method of synthesis. Thus "amino acid", or "peptide residue", as used herein means both naturally occurring and synthetic amino acids. For example, homo-phenylalanine, citrulline, and noreleucine are considered amino acids for the purposes of the invention. "Amino acid" also includes imino acid residues such as proline and hydroxyproline. In addition, any amino acid representing a component of the variant proteins of the present invention can be replaced by the same amino acid but of the opposite chirality. Thus, any amino acid naturally occurring in the L-configuration (which may also be referred to as the R or S, depending upon the structure of the chemical entity) may be replaced with an amino acid of the same chemical structural type, but of the opposite chirality, generally referred to as the D- amino acid but which can additionally be referred to as the R- or the S-, depending upon its composition and chemical configuration. Such derivatives have the property of greatly increased stability, and therefore are advantageous in the formulation of compounds which may have longer in vivo half lives, when administered by oral, intravenous, intramuscular, intraperitoneal, topical, rectal, intraocular, or other routes. In the preferred embodiment, the amino acids are in the (S) or L-configuration. If non-naturally occurring side chains are used, non-amino acid substituents may be used, for example to prevent or retard in vivo degradations. Proteins including non-naturally occurring amino acids may be synthesized or in some cases, made recombinantly; see van Hest et al., FEBS Lett 428:(1-2) 68-70 May 22 1998 and Tang et al., Abstr. Pap Am. Chem. S218:U138-U138 Part 2 August 22, 1999, both of which are expressly incorporated by reference herein.

Aromatic amino acids may be replaced with D- or L-naphthylalanine, D- or L-Phenylglycine, D- or L-2-thienylalanine, D- or L-1-, 2-, 3- or 4-pyrenylalanine, D- or L-3-thienylalanine, D- or L-(2-pyridinyl)-alanine, D- or L-(3-pyridinyl)-alanine, D- or L-(2-pyrazinyl)-alanine, D- or L-(4-isopropyl)-phenylglycine, D-(trifluoromethyl)-phenylglycine, D-(trifluoromethyl)-phenylalanine, D-p-fluorophenylalanine, D- or L-p-biphenylphenylalanine, D- or L-p-methoxybiphenylphenylalanine, D- or L-2-indole(alkyl)alanines, and D- or L-alkylalanines where alkyl may be substituted or unsubstituted methyl, ethyl, propyl, hexyl, butyl, pentyl, isopropyl, iso-butyl, sec-isotyl, iso-pentyl, non-acidic amino acids, of C1-C20. Acidic amino acids can be substituted with non-carboxylate amino acids while maintaining a negative charge, and derivatives or analogs thereof, such as the non-limiting examples of (phosphono)alanine, glycine, leucine, isoleucine, threonine, or serine; or sulfated (e.g., -SO₃H) threonine, serine, tyrosine. Other substitutions may include unnatural hydroxylated amino acids may be made by combining "alkyl" with any natural amino acid. The term "alkyl" as used herein refers to a branched or unbranched saturated hydrocarbon group of 1 to 24 carbon atoms, such as methyl, ethyl, n-propyl, isopropyl, n-butyl, isobutyl, t-butyl, octyl, decyl, tetradecyl, hexadecyl, eicosyl, tetracosyl and the like.

WO 02/090300

PCT/US02/14358

Alkyl includes heteroalkyl, with atoms of nitrogen, oxygen and sulfur. Preferred alkyl groups herein contain 1 to 12 carbon atoms. Basic amino acids may be substituted with alkyl groups at any position of the naturally occurring amino acids lysine, arginine, ornithine, citrulline, or (guanidino)-acetic acid, or other (guanidino)alkyl-acetic acids, where "alkyl" is defined as above. Nitrile derivatives (e.g., containing the CN-moiety in place of COOH) may also be substituted for asparagine or glutamine, and methionine sulfoxide may be substituted for methionine. Methods of preparation of such peptide derivatives are well known to one skilled in the art.

In addition, any amide linkage in any of the variant polypeptides can be replaced by a ketomethylene moiety. Such derivatives are expected to have the property of increased stability to degradation by enzymes, and therefore possess advantages for the formulation of compounds which may have increased in vivo half lives, as administered by oral, intravenous, intramuscular, intraperitoneal, topical, rectal, intraocular, or other routes. Additional amino acid modifications of amino acids of variant polypeptides of to the present invention may include the following: Cysteiny residues may be reacted with alpha-haloacetates (and corresponding amines), such as 2-chloroacetic acid or chloroacetamide, to give carboxymethyl or carboxyamidomethyl derivatives. Cysteiny residues may also be derivatized by reaction with compounds such as bromotrifluoroacetone, alpha-bromo-beta-(5-imidozoyl)propionic acid, chloroacetyl phosphate, N-alkylmaleimides, 3-nitro-2-pyridyl disulfide, methyl 2-pyridyl disulfide, p-chloromercuribenzoate, 2-chloromercuri-4-nitrophenol, or chloro-7-nitrobenzo-2-oxa-1,3-diazole. Histidyl residues may be derivatized by reaction with compounds such as diethylprocarbonate e.g., at pH 5.5-7.0 because this agent is relatively specific for the histidyl side chain, and para-bromophenacyl bromide may also be used; e.g., where the reaction is preferably performed in 0.1M sodium cacodylate at pH 6.0. Lysiny and amino terminal residues may be reacted with compounds such as succinic or other carboxylic acid anhydrides. Derivatization with these agents is expected to have the effect of reversing the charge of the lysiny residues. Other suitable reagents for derivatizing alpha-amino-containing residues include compounds such as imidoesters/e.g., as methyl picolinimate; pyridoxal phosphate; pyridoxal; chloroborohydride; trinitrobenzenesulfonic acid; O-methylisourea; 2,4 pentanedione; and transaminase-catalyzed reaction with glyoxylate. Arginy residues may be modified by reaction with one or several conventional reagents, among them phenylglyoxal, 2,3-butanedione, 1,2-cyclohexanedione, and ninhydrin according to known method steps. Derivatization of arginine residues requires that the reaction be performed in alkaline conditions because of the high pKa of the guanidine functional group. Furthermore, these reagents may react with the groups of lysine as well as the arginine epsilon-amino group. The specific modification of tyrosyl residues per se is well-known, such as for introducing spectral labels into tyrosyl residues by reaction with aromatic diazonium compounds or tetranitromethane. Arginy residues may be modified by reaction with one or several conventional reagents, among them phenylglyoxal, 2,3-butanedione, 1,2-cyclohexanedione, and ninhydrin according to known method steps. Derivatization of arginine residues requires that the reaction be performed in alkaline conditions because of the high pKa of the guanidine functional group. Furthermore, these reagents may react with the groups of lysine as well as the arginine epsilon-amino group. The specific modification of tyrosyl residues per se is well-known, such as for introducing spectral labels into tyrosyl residues by reaction with aromatic diazonium compounds or tetranitromethane. N-acetylimidazol and tetranitromethane may be used to form O-acetyl tyrosyl species and 3-nitro derivatives, respectively. Carboxyl side groups (aspartyl or glutamyl) may be selectively modified by reaction with carbodiimides (R'-N-C-N-R') such as 1-cyclohexyl-3-(2-morpholinyl)- (4-ethyl) carbodiimide or 1-ethyl-3-(4-azonia-4,4-dimethylpentyl) carbodiimide.

WO 02/090300

PCT/US02/14358

Furthermore aspartyl and glutamyl residues may be converted to asparaginy and glutaminy residues by reaction with ammonium ions. Glutaminy and asparaginy residues may be frequently deamidated to the corresponding glutamyl and aspartyl residues. Alternatively, these residues may be deamidated under mildly acidic conditions. Either form of these residues falls within the scope of the present invention.

The target or scaffold protein may be any protein for which a three dimensional structure is known or can be generated; that is, for which there are three dimensional coordinates for each atom of the protein. Generally this can be determined using X-ray crystallographic techniques, NMR techniques, de novo modeling, homology modeling, etc. In general, if X-ray structures are used, structures at 2 resolution or better are preferred, but not required.

The target or scaffold proteins of the present invention may be from prokaryotes and eukaryotes, such as bacteria (including extremeophiles such as the archebacteria), fungi, insects, fish, plants, and mammals. Suitable mammals include, but are not limited to, rodents (rats, mice, hamsters, guinea pigs, etc.), primates, farm animals (including sheep, goats, pigs, cows, horses, etc) and in the most preferred embodiment, from humans.

Thus, by "target protein" or "scaffold protein" herein is meant a protein for which a variant protein or a library of variant proteins, preferably with altered cofactor specificity is desired. As will be appreciated by those in the art, any number of target proteins find use in the present invention. Specifically included within the definition of "protein" are fragments and domains of known proteins, including functional domains such as enzymatic domains, binding domains, etc., and smaller fragments, such as turns, loops, etc. That is, portions of proteins may be used as well. In addition, "protein" as used herein includes proteins, oligopeptides and peptides. In addition, protein variants, i.e. non-naturally occurring protein analog structures, may be used.

Suitable proteins include, but are not limited to, industrial, pharmaceutical, and agricultural proteins. Suitable classes of enzymes include, but are not limited to, reductases, hydrolases such as proteases, carbohydrases, lipases; isomerases such as racemases, epimerases, tautomerases, or mutases; transferases, kinases, oxidoreductases, dehydrogenases, and phosphatases. Suitable enzymes are listed in the Swiss-Prot enzyme database. Suitable protein backbones include, but are not limited to, all of those found in the protein data base compiled and serviced by the Research Collaboratory for Structural Bioinformatics (RCSB, formerly the Brookhaven National Lab).

Specifically, preferred target protein include reductases, such as thioredoxin reductase (US Pub. No. 2002/0037303), 2,5-diketo-D-gluconic acid reductase (Banta, S, et al., (2002) *Protein Eng.*, 15: 131-140; WO 02/22527; WO 02/29019), glutathione reductase (Mittl, PR, et al. (1993) *J. Mol. Biol.*, 231: 191-5; Mittl & Schulz, (1994) *Protein Sci.*, 3: 799-809; Mittl, PR, et al., (1994) *Protein Sci.*, 3: 1504-

WO 02/090300

PCT/US02/14358

14), the alkyl hydroperoxide reductase system (Wood, ZA, et al., (2001), *Biochemistry*, 40: 3900-3911), thioredoxin reductase-like proteins (Reynolds, CM, et al., (2002) *Biochemistry*, 41: 1990-2001)

Accordingly, the present invention is directed to computational processing methods for altering the cofactor specificity of the target protein. Once a set of coordinates for a target protein or scaffold protein is imported, a protein design cycle is implemented to generate a set of variable protein sequences with altered affinity for a desired receptor. By "protein design cycle" herein is meant any one of a number of protein design algorithms that can be used to produce a sequence or sequence including but not limited to Protein Design Automation™ (PDA™), sequence prediction algorithm (SPA), various force field calculations, etc. See U.S. Patent Nos. 6,188,965 and 6,296,312, U.S.S.N.s 09/419,351, 09/762,004, 09/927,79, 09/877,695; Raha, K., et al. (2000) *Protein Sci.*, 9:1106-1119, U.S.S.N. 09/877,695, filed June 8, 2001, entitled "Apparatus and Method for Designing Proteins and Protein Libraries; U.S.S.N.s 09/927,790, 60/352,103, and 60/351,937, all of which are expressly incorporated herein by reference in their entirety.

In a preferred embodiment, the methods of the invention involve starting with a target protein and use computational processing to generate a candidate or variant protein or a set of primary sequences. In a preferred embodiment, sequence based methods are used. Alternatively, structure based methods, such as PDA™, described in detail below, are used. Other models for assessing the relative energies of sequences with high precision include Warshel, Computer Modelling of Chemical Reactions in Enzymes and Solutions, Wiley & Sons, New York, (1991), hereby expressly incorporated by reference.

Similarly, molecular dynamics calculations can be used to computationally screen sequences by individually calculating mutant sequence scores and compiling a rank ordered list.

In a preferred embodiment, residue pair potentials can be used to score sequences (Miyazawa et al., *Macromolecules* 18(3):534-552 (1985), expressly incorporated by reference) during computational screening.

In a preferred embodiment, sequence profile scores (Bowie et al., *Science* 253(5016):164-70 (1991), incorporated by reference) and/or potentials of mean force (Hendlich et al., *J. Mol. Biol.* 216(1):167-180 (1990), also incorporated by reference) can also be calculated to score sequences. These methods assess the match between a sequence and a 3D protein structure and hence can act to screen for fidelity to the protein structure. By using different scoring functions to rank sequences, different regions of sequence space can be sampled in the computational screen.

Furthermore, scoring functions can be used to screen for sequences that would create metal or co-factor binding sites in the protein (Hellinga, *Fold Des.* 3(1): R1-8 (1998), hereby expressly

WO 02/090300

PCT/US02/14358

incorporated by reference). Similarly, scoring functions can be used to screen for sequences that would create disulfide bonds in the protein. These potentials attempt to specifically modify a protein structure to introduce a new structural motif.

- 5 In a preferred embodiment, sequence and/or structural alignment programs can be used to generate the variant proteins of the invention. As is known in the art, there are a number of sequence-based alignment programs; including for example, Smith-Waterman searches, Needleman-Wunsch, Double Affine Smith-Waterman, frame search, Gribskov/GCG profile search, Gribskov/GCG profile scan, profile frame search, Bucher generalized profiles, Hidden Markov models, Hframe, Double Frame, 10 Blast, Psi-Blast, Clustal, and GeneWise.

- The source of the sequences can vary widely, and include taking sequences from one or more of the known databases, including, but not limited to, SCOP (Hubbard, et al., Nucleic Acids Res 27(1):254-256. (1999)); PFAM (Bateman, et al., Nucleic Acids Res 27(1):260-262. (1999)); VAST (Gibrat, et al., Curr Opin Struct Biol 6(3):377-385. (1996)); CATH (Orengo, et al., Structure 5(8):1093-1108. (1997)); 15 Phd Predictor (<http://www.embl-heidelberg.de/predictprotein/predictprotein.html>); Prosite (Hofmann, et al., Nucleic Acids Res 27(1):215-219. (1999)); PIR (<http://www.mips.biochem.mpg.de/proj/protseqdb/>); GenBank (<http://www.ncbi.nlm.nih.gov/>); PDB (www.rcsb.org) and BIND (Bader, et al., Nucleic Acids Res 29(1):242-245. (2001)).
- 20 In addition, sequences from these databases can be subjected to contiguous analysis or gene prediction; see Wheeler, et al., Nucleic Acids Res 28(1):10-14. (2000) and Burge and Karlin, J Mol Biol 268(1):78-94. (1997).

- As is known in the art, there are a number of sequence alignment methodologies that can be used. 25 For example, sequence homology based alignment methods can be used to create sequence alignments of proteins related to the target structure (Altschul et al., J. Mol. Biol. 215(3):403-410 (1990), Altschul et al., Nucleic Acids Res. 25:3389-3402 (1997), both incorporated by reference). These sequence alignments are then examined to determine the observed sequence variations. These sequence variations are tabulated to define a set of variant proteins.

- 30 Sequence based alignments can be used in a variety of ways. For example, a number of related proteins can be aligned, as is known in the art, and the "variable" and "conserved" residues defined; that is, the residues that vary or remain identical between the family members can be defined. These results can be used to generate a probability table, as outlined below. Similarly, these sequence variations can be tabulated and a secondary library defined from them as defined below. 35 Alternatively, the allowed sequence variations can be used to define the amino acids considered at each position during the computational screening. Another variation is to bias the score for amino acids that occur in the sequence alignment, thereby increasing the likelihood that they are found during computational screening but still allowing consideration of other amino acids. This bias would

WO 02/090300

PCT/US02/14358

result in a focused library of variant proteins but would not eliminate from consideration amino acids not found in the alignment. In addition, a number of other types of bias may be introduced. For example, diversity may be forced; that is, a "conserved" residue is chosen and altered to force diversity on the protein and thus sample a greater portion of the sequence space. Alternatively, the positions of high variability between family members (i.e. low conservation) can be randomized, either using all or a subset of amino acids. Similarly, outlier residues, either positional outliers or side chain outliers, may be eliminated.

Similarly, structural alignment of structurally related proteins can be done to generate sequence alignments. There are a wide variety of such structural alignment programs known. See for example VAST from the NCBI (<http://www.ncbi.nlm.nih.gov/80/Structure/VAST/vast.shtml>); SSAP (Orengo and Taylor, *Methods Enzymol* 266(617-635 (1996)) SARF2 (Alexandrov, *Protein Eng* 9(9):727-732. (1996)) CE (Shindyalov and Bourne, *Protein Eng* 11(9):739-747. (1998)); (Orengo et al., *Structure* 5(8):1093-108 (1997); Dali (Holm et al., *Nucleic Acid Res.* 26(1):316-9 (1998), all of which are incorporated by reference). These sequence alignments can then be examined to determine the observed sequence variations. Libraries can be generated by predicting secondary structure from sequence, and then selecting sequences that are compatible with the predicted secondary structure. There are a number of secondary structure prediction methods such as helix-coil transition theory (Munoz and Serrano, *Biopolymers* 41:495, 1997), neural networks, local structure alignment and others (e.g., see in Selbig et al., *Bioinformatics* 15:1039-46, 1999).

Similarly, as outlined above, other computational methods are known, including, but not limited to, sequence profiling [Bowie and Eisenberg, *Science* 253(5016):164-70, (1991)], rotamer library selections [Dahiyat and Mayo, *Protein Sci.* 5(6):896-903 (1996); Dahiyat and Mayo, *Science* 278(5335):82-7 (1997); Desjarlais and Handel, *Protein Science* 4:2006-2018 (1995); Harbury et al, *Proc. Natl. Acad. Sci. U.S.A.* 92(18):8408-8412 (1995); Kono et al., *Proteins: Structure, Function and Genetics* 19:244-255 (1994); Hellinga and Richards, *Proc. Natl. Acad. Sci. U.S.A.* 91:5803-5807 (1994)]; and residue pair potentials [Jones, *Protein Science* 3: 567-574, (1994)]; PROSA [Heindlich et al., *J. Mol. Biol.* 216:167-180 (1990)]; THREADER [Jones et al., *Nature* 358:86-89 (1992)], and other inverse folding methods such as those described by Simons et al. [*Proteins*, 34:535-543, (1999)], Levitt and Gerstein [*Proc. Natl. Acad. Sci. U.S.A.*, 95:5913-5920, (1998)], Godzik and Skolnick [*Proc. Natl. Acad. Sci. U.S.A.*, 89:12098-102, (1992)], Godzik et al. [*J. Mol. Biol.* 227:227-38, (1992)] and two profile methods [Gribkov et al. *Proc. Natl. Acad. Sci. U.S.A.* 84:4365-4358 (1987) and Fischer and Eisenberg, *Protein Sci.* 5:947-955 (1996), Rice and Eisenberg *J. Mol. Biol.* 267:1026-1038(1997)], all of which are expressly incorporated by reference.

In addition, other computational methods such as those described by Koehl and Levitt (*J. Mol. Biol.* 293:1161-1181 (1999); *J. Mol. Biol.* 293:1183-1193 (1999); expressly incorporated by reference) can be used to create a variant library that can optionally then be used to generate a smaller secondary

WO 02/090300

PCT/US02/14358

library for use in experimental screening for improved properties and function. In addition, there are computational methods based on force-field calculations such as SCMF that can be used as well for SCMF, see Delarue et al. *Pac. Symp. Biocomput.* 109-21 (1997); Koehl et al., *J. Mol. Biol.* 239:249-75 (1994); Koehl et al., *Nat. Struct. Biol.* 2:163-70 (1995); Koehl et al., *Curr. Opin. Struct. Biol.* 6:222-6 (1996); Koehl et al., *J. Mol. Biol.* 293:1183-93 (1999); Koehl et al., *J. Mol. Biol.* 293:1161-81 (1999); Lee J., *Mol. Biol.* 236:918-39 (1994); and Vasquez *Biopolymers* 36:53-70 (1995); all of which are expressly incorporated by reference. Other forcefield calculations that can be used to optimize the conformation of a sequence within a computational method, or to generate de novo optimized sequences as outlined herein include, but are not limited to, OPLS-AA [Jorgensen et al., *J. Am. Chem. Soc.* 118:11225-11236 (1996); Jorgensen, W.L.; BOSS, Version 4.1; Yale University: New Haven, CT (1999)]; OPLS [Jorgensen et al., *J. Am. Chem. Soc.* 110:1657ff (1988); Jorgensen et al., *J. Am. Chem. Soc.* 112:4768ff (1990)]; UNRES (United Residue Forcefield; Liwo et al., *Protein Science* 2:1697-1714 (1993); Liwo et al., *Protein Science* 2:1715-1731 (1993); Liwo et al., *J. Comp. Chem.* 18:849-873 (1997); Liwo et al., *J. Comp. Chem.* 18:874-884 (1997); Liwo et al., *J. Comp. Chem.* 19:259-276 (1998); Forcefield for Protein Structure Prediction (Liwo et al., *Proc. Natl. Acad. Sci. U.S.A.* 96:5482-5485 (1999)]; ECEPP/3 [Liwo et al., *J. Protein Chem.* 13(4):375-80 (1994)]; AMBER 1.1 force field (Weiner et al., *J. Am. Chem. Soc.* 106:765_784); AMBER 3.0 force field [U.C. Singh et al., *Proc. Natl. Acad. Sci. U.S.A.* 82:755-759 (1985)]; CHARMM and CHARMM22 (Brooks et al., *J. Comp. Chem.* 4:187-217); cvff3.0 [Dauber-Osguthorpe et al., *Proteins: Structure, Function and Genetics*, 4:31-47 (1988)]; cff91 (Maple et al., *J. Comp. Chem.* 15:162-182); also, the DISCOVER (cvff and cff91) and AMBER force-fields are used in the INSIGHT molecular modeling package (Biosym/MSI, San Diego California) and HARM is used in the QUANTA molecular modeling package (Biosym/MSI, San Diego California), all of which are expressly incorporated by reference. In fact, as is outlined below, these force-field methods may be used to generate the variant TR library directly; these methods can be used to generate a probability table from which an additional library is directly generated.

In a preferred embodiment, Protein Design Automation™ (PDA™) is used to generate a variable protein sequence comprising a defined energy state for each amino acid position as is described in U.S. Patent Nos. 6,188,965 and 6,296,312, all of which are expressly incorporated herein by reference. Briefly, PDA™ can be described as follows. A known protein structure is used as the starting point. The residues to be optimized are then identified, which may be the entire sequence or subset(s) thereof. The side chains of any positions to be varied are then removed. The resulting structure consisting of the protein backbone and the remaining sidechains is called the template. Each variable residue position is then preferably classified as a core residue, a surface residue, or a boundary residue; each classification defines a subset of possible amino acid residues for the position (for example, core residues generally will be selected from the set of hydrophobic residues, surface residues generally will be selected from the hydrophilic residues, and boundary residues may be either). Each amino acid can be represented by a discrete set of all allowed conformers of each side

WO 02/090300

PCT/US02/14358

chain, called rotamers. Thus, to arrive at an optimal sequence for a backbone, all possible sequences of rotamers must be screened, where each backbone position can be occupied either by each amino acid in all its possible rotameric states, or a subset of amino acids, and thus a subset of rotamers.

5

Two sets of interactions are then calculated for each rotamer at every position: the interaction of the rotamer side chain with all or part of the backbone (the "singles" energy, also called the rotamer/template or rotamer/backbone energy), and the interaction of the rotamer side chain with all other possible rotamers at every other position or a subset of the other positions (the "doubles" energy, also called the rotamer/rotamer energy). The energy of each of these interactions is calculated through the use of a variety of scoring functions, which include the energy of van der Waal's forces, the energy of hydrogen bonding, the energy of secondary structure propensity, the energy of surface area solvation and the electrostatics. Thus, the total energy of each rotamer interaction, both with the backbone and other rotamers, is calculated, and stored in a matrix form.

15

The discrete nature of rotamer sets allows a simple calculation of the number of rotamer sequences to be tested. A backbone of length n with m possible rotamers per position will have m^n possible rotamer sequences, a number which grows exponentially with sequence length and renders the calculations either unwieldy or impossible in real time. Accordingly, to solve this combinatorial search problem, a "Dead End Elimination" (DEE) calculation is performed. The DEE calculation is based on the fact that if the worst total interaction of a first rotamer is still better than the best total interaction of a second rotamer, then the second rotamer cannot be part of the global optimum solution. Since the energies of all rotamers have already been calculated, the DEE approach only requires sums over the sequence length to test and eliminate rotamers, which speeds up the calculations considerably. DEE can be rerun comparing pairs of rotamers, or combinations of rotamers, which will eventually result in the determination of a single sequence which represents the global optimum energy.

20

25

Once the global solution has been found, a Monte Carlo search may be done to generate a rank-ordered list of sequences in the neighborhood of the DEE solution. Starting at the DEE solution, random positions are changed to other rotamers, and the new sequence energy is calculated. If the new sequence meets the criteria for acceptance, it is used as a starting point for another jump. After a predetermined number of jumps, a rank-ordered list of sequences is generated. Monte Carlo searching is a sampling technique to explore sequence space around the global minimum or to find new local minima distant in sequence space. As is more additionally outlined below, there are other sampling techniques that can be used, including Boltzman sampling, genetic algorithm techniques and simulated annealing. In addition, for all the sampling techniques, the kinds of jumps allowed can be altered (e.g. random jumps to random residues, biased jumps (to or away from wild-type, for example), jumps to biased residues (to or away from similar residues, for example), etc.). Similarly,

30

35

WO 02/090300

PCT/US02/14358

for all the sampling techniques, the acceptance criteria of whether a sampling jump is accepted can be altered.

5 As outlined in U.S.S.N. 09/127,926, the protein backbone (comprising (for a naturally occurring protein) the nitrogen, the carbonyl carbon, the α -carbon, and the carbonyl oxygen, along with the direction of the vector from the α -carbon to the β -carbon) may be altered prior to the computational analysis, by varying a set of parameters called supersecondary structure parameters.

10 Once a protein structure backbone is generated (with alterations, as outlined above) and input into the computer, explicit hydrogens are added if not included within the structure (for example, if the structure was generated by X-ray crystallography, hydrogens must be added). After hydrogen addition, energy minimization of the structure is run, to relax the hydrogens as well as the other atoms, bond angles and bond lengths. In a preferred embodiment, this is done by doing a number of steps of conjugate gradient minimization (Mayo *et al.*, J. Phys. Chem. 94:8897 (1990)) of atomic coordinate positions to minimize the Dreiding force field with no electrostatics. Generally from about 15 10 to about 250 steps is preferred, with about 50 being most preferred.

The protein backbone structure contains at least one variable residue position. As is known in the art, the residues, or amino acids, of proteins are generally sequentially numbered starting with the N-terminus of the protein. Thus a protein having a methionine at it's N-terminus is said to have a 20 methionine at residue or amino acid position 1, with the next residues as 2, 3, 4, etc. At each position, the wild type (i.e. naturally occurring) protein may have one of at least 20 amino acids, in any number of rotamers. By "variable residue position" herein is meant an amino acid position of the protein to be designed that is not fixed in the design method as a specific residue or rotamer, 25 generally the wild-type residue or rotamer.

In a preferred embodiment, all of the residue positions of the protein are variable. That is, every amino acid side chain may be altered in the methods of the present invention. This is particularly desirable for smaller proteins, although the present methods allow the design of larger proteins as 30 well. While there is no theoretical limit to the length of the protein that may be designed this way, there is a practical computational limit.

In an alternate preferred embodiment, only some of the residue positions of the protein are variable, and the remainder are "fixed", that is, they are identified in the three dimensional structure as being in 35 a set conformation. In some embodiments, a fixed position is left in its original conformation (which may or may not correlate to a specific rotamer of the rotamer library being used). Alternatively, residues may be fixed as a non-wild type residue; for example, when known site-directed mutagenesis techniques have shown that a particular residue is desirable (for example, to eliminate a proteolytic site or alter the substrate specificity of an enzyme), the residue may be fixed as a particular

WO 02/090300

PCT/US02/14358

amino acid. Alternatively, the methods of the present invention may be used to evaluate mutations de novo, as is discussed below. In an alternate preferred embodiment, a fixed position may be "floated"; the amino acid at that position is fixed, but different rotamers of that amino acid are tested. In this embodiment, the variable residues may be at least one, or anywhere from 0.1% to 99.9% of the total number of residues. Thus, for example, it may be possible to change only a few (or one) residues, or most of the residues, with all possibilities in between.

In a preferred embodiment, residues which can be fixed include, but are not limited to, structurally or biologically functional residues; alternatively, biologically functional residues may specifically not be fixed. For example, residues which are known to be important for biological activity, such as the residues which form the active site of an enzyme, the substrate binding site of an enzyme, the binding site for a binding partner (ligand/receptor, antigen/antibody, etc.), phosphorylation or glycosylation sites which are crucial to biological function, or structurally important residues, such as disulfide bridges, metal binding sites, critical hydrogen bonding residues, residues critical for backbone conformation such as proline or glycine, residues critical for packing interactions, etc. may all be fixed in a conformation or as a single rotamer, or "floated".

Similarly, residues which may be chosen as variable residues may be those that confer undesirable biological attributes, such as susceptibility to proteolytic degradation, dimerization or aggregation sites, glycosylation sites which may lead to immune responses, unwanted binding activity, unwanted allostery, undesirable enzyme activity but with a preservation of binding, etc.

In a preferred embodiment, each variable position is classified as either a core, surface or boundary residue position, although in some cases, as explained below, the variable position may be set to glycine to minimize backbone strain. In addition, as outlined herein, residues need not be classified, they can be chosen as variable and any set of amino acids may be used. Any combination of core, surface and boundary positions can be utilized: core, surface and boundary residues; core and surface residues; core and boundary residues, and surface and boundary residues, as well as core residues alone, surface residues alone, or boundary residues alone.

Classification of residue positions as core, surface or boundary may be done in several ways, as will be appreciated by those of skill in the art. In a preferred embodiment, the classification is done via a visual scan of the original protein backbone structure, including the side chains, and assigning a classification based on a subjective evaluation of one skilled in the art of protein modeling.

Alternatively, a preferred embodiment utilizes an assessment of the orientation of the C α -C β vectors relative to a solvent accessible surface computed using only the template C α atoms, as outlined in U.S.S.N.s 60/061,097, 60/043,464, 60/054,678, 09/127,926 and PCT US98/07254. Alternatively, a surface area calculation can be done.

WO 02/090300

PCT/US02/14358

Once each variable position is optionally classified as either core, surface or boundary, a set of amino acid side chains, and thus a set of rotamers, is assigned to each position. That is, the set of possible amino acid side chains that the program will allow to be considered at any particular position is chosen. Subsequently, once the possible amino acid side chains are chosen, the set of rotamers that will be evaluated at a particular position can be determined. Thus, a core residue will generally be selected from the group of hydrophobic residues consisting of alanine, valine, isoleucine, leucine, phenylalanine, tyrosine, tryptophan, and methionine (in some embodiments, when the scaling factor of the van der Waals scoring function, described below, is low, methionine is removed from the set), and the rotamer set for each core position potentially includes rotamers for these eight amino acid side chains (all the rotamers if a backbone independent library is used, and subsets if a rotamer dependent backbone is used). Similarly, surface positions are generally selected from the group of hydrophilic residues consisting of alanine, serine, threonine, aspartic acid, asparagine, glutamine, glutamic acid, arginine, lysine and histidine. The rotamer set for each surface position thus includes rotamers for these ten residues. Finally, boundary positions are generally chosen from alanine, serine, threonine, aspartic acid, asparagine, glutamine, glutamic acid, arginine, lysine histidine, valine, isoleucine, leucine, phenylalanine, tyrosine, tryptophan, and methionine. The rotamer set for each boundary position thus potentially includes every rotamer for these seventeen residues (assuming cysteine, glycine and proline are not used, although they can be). Additionally, in some preferred embodiments, a set of 18 naturally occurring amino acids (all except cysteine and proline, which are known to be particularly disruptive) are used.

Thus, as will be appreciated by those in the art, there is a computational benefit to classifying the residue positions, as it decreases the number of calculations. It should also be noted that there may be situations where the sets of core, boundary and surface residues are altered from those described above; for example, under some circumstances, one or more amino acids is either added or subtracted from the set of allowed amino acids. For example, some proteins that dimerize or multimerize, or have ligand binding sites, may contain hydrophobic surface residues, etc. In addition, residues that do not allow helix "capping" or the favorable interaction with an α -helix dipole may be subtracted from a set of allowed residues. This modification of amino acid groups is done on a residue by residue basis.

In a preferred embodiment, proline, cysteine and glycine are not included in the list of possible amino acid side chains, and thus the rotamers for these side chains are not used. However, in a preferred embodiment, when the variable residue position has a Φ angle (that is, the dihedral angle defined by 1) the carbonyl carbon of the preceding amino acid; 2) the nitrogen atom of the current residue; 3) the α -carbon of the current residue; and 4) the carbonyl carbon of the current residue) greater than 0° , the position is set to glycine to minimize backbone strain.

WO 02/090300

PCT/US02/14358

Once the group of potential rotamers is assigned for each variable residue position, processing proceeds as outlined in U.S.S.N. 09/127,926 and PCT US98/07254. This processing step entails analyzing interactions of the rotamers with each other and with the protein backbone to generate optimized protein sequences. Simplistically, the processing initially comprises the use of a number of scoring functions to calculate energies of interactions of the rotamers, either to the backbone itself or other rotamers. Preferred PDA scoring functions include, but are not limited to, a Van der Waals potential scoring function, a hydrogen bond potential scoring function, an atomic solvation scoring function, a secondary structure propensity scoring function and an electrostatic scoring function. As is further described below, at least one scoring function is used to score each position, although the scoring functions may differ depending on the position classification or other considerations, like favorable interaction with an α -helix dipole. As outlined below, the total energy which is used in the calculations is the sum of the energy of each scoring function used at a particular position, as is generally shown in Equation 1:

Equation 1

$$E_{\text{total}} = nE_{\text{vdw}} + nE_{\text{as}} + nE_{\text{h-bonding}} + nE_{\text{ss}} + nE_{\text{elec}}$$

In Equation 1, the total energy is the sum of the energy of the van der Waals potential (E_{vdw}), the energy of atomic solvation (E_{as}), the energy of hydrogen bonding ($E_{\text{h-bonding}}$), the energy of secondary structure (E_{ss}) and the energy of electrostatic interaction (E_{elec}). The term n is either 0 or 1, depending on whether the term is to be considered for the particular residue position.

As outlined in U.S.S.N.s 60/061,097, 60/043,464, 60/054,678, 09/127,926 and PCT US98/07254, any combination of these scoring functions, either alone or in combination, may be used. Once the scoring functions to be used are identified for each variable position, the preferred first step in the computational analysis comprises the determination of the interaction of each possible rotamer with all or part of the remainder of the protein. That is, the energy of interaction, as measured by one or more of the scoring functions, of each possible rotamer at each variable residue position with either the backbone or other rotamers, is calculated. In a preferred embodiment, the interaction of each rotamer with the entire remainder of the protein, i.e. both the entire template and all other rotamers, is done. However, as outlined above, it is possible to only model a portion of a protein, for example a domain of a larger protein, and thus in some cases, not all of the protein need be considered. The term "portion", as used herein, with regard to a protein refers to a fragment of that protein. This fragment may range in size from 10 amino acid residues to the entire amino acid sequence minus one amino acid. Accordingly, the term "portion", as used herein, with regard to a nucleic refers to a fragment of that nucleic acid. This fragment may range in size from 10 nucleotides to the entire nucleic acid sequence minus one nucleotide.

In a preferred embodiment, the first step of the computational processing is done by calculating two sets of interactions for each rotamer at every position: the interaction of the rotamer side chain with the template or backbone (the "singles" energy), and the interaction of the rotamer side chain with all

WO 02/090300

PCT/US02/14358

other possible rotamers at every other position (the "doubles" energy), whether that position is varied or floated. It should be understood that the backbone in this case includes both the atoms of the protein structure backbone, as well as the atoms of any fixed residues, wherein the fixed residues are defined as a particular conformation of an amino acid.

5

Thus, "singles" (rotamer/template) energies are calculated for the interaction of every possible rotamer at every variable residue position with the backbone, using some or all of the scoring functions. Thus, for the hydrogen bonding scoring function, every hydrogen bonding atom of the rotamer and every hydrogen bonding atom of the backbone is evaluated, and the E_{HB} is calculated for each possible rotamer at every variable position. Similarly, for the van der Waals scoring function, every atom of the rotamer is compared to every atom of the template (generally excluding the backbone atoms of its own residue), and the E_{vdW} is calculated for each possible rotamer at every variable residue position. In addition, generally no van der Waals energy is calculated if the atoms are connected by three bonds or less. For the atomic solvation scoring function, the surface of the rotamer is measured against the surface of the template, and the E_{sa} for each possible rotamer at every variable residue position is calculated. The secondary structure propensity scoring function is also considered as a singles energy, and thus the total singles energy may contain an E_{ss} term. As will be appreciated by those in the art, many of these energy terms will be close to zero, depending on the physical distance between the rotamer and the template position; that is, the farther apart the two moieties, the lower the energy.

20

For the calculation of "doubles" energy (rotamer/rotamer), the interaction energy of each possible rotamer is compared with every possible rotamer at all other variable residue positions. Thus, "doubles" energies are calculated for the interaction of every possible rotamer at every variable residue position with every possible rotamer at every other variable residue position, using some or all of the scoring functions. Thus, for the hydrogen bonding scoring function, every hydrogen bonding atom of the first rotamer and every hydrogen bonding atom of every possible second rotamer is evaluated, and the E_{HB} is calculated for each possible rotamer pair for any two variable positions. Similarly, for the van der Waals scoring function, every atom of the first rotamer is compared to every atom of every possible second rotamer, and the E_{vdW} is calculated for each possible rotamer pair at every two variable residue positions. For the atomic solvation scoring function, the surface of the first rotamer is measured against the surface of every possible second rotamer, and the E_{sa} for each possible rotamer pair at every two variable residue positions is calculated. The secondary structure propensity scoring function need not be run as a "doubles" energy, as it is considered as a component of the "singles" energy. As will be appreciated by those in the art, many of these double energy terms will be close to zero, depending on the physical distance between the first rotamer and the second rotamer; that is, the farther apart the two moieties, the lower the energy.

25

30

35

WO 02/090300

PCT/US02/14358

In a preferred embodiment, a sequence prediction algorithm (SPA) is used to generate a variable protein sequence comprising a defined energy state for each amino acid position as is described in Raha, K., et al. (2000) *Protein Sci.*, 9:1106-1119, U.S.S.N. 09/877,695, filed June 8, 2001, entitled "Apparatus and Method for Designing Proteins and Protein Libraries"; both of which are expressly incorporated herein by reference.

- In a preferred embodiment, force field calculations such as SCMF can be used generate a variable protein sequence comprising a defined energy state for each amino acid position. For SCMF, see Delarue et al., *Pac. Symp. Biocomput.* 109-21 (1997); Koehl et al., *J. Mol. Biol.* 239:249 (1994); Koehl et al., *Nat. Struct. Biol.* 2:163 (1995); Koehl et al., *Curr. Opin. Struct. Biol.* 6:222 (1996); Koehl et al., *J. Mol. Bio.* 293:1183 (1999); Koehl et al., *J. Mol. Biol.* 293:1161 (1999); Lee J., *Mol. Biol.* 236:918 (1994); and Vasquez *Biopolymers* 36:53-70 (1995); all of which are expressly incorporated by reference. Other force field calculations that can be used to optimize the conformation of a sequence within a computational method, or to generate de novo optimized sequences as outlined herein include, but are not limited to, OPLS_AA (Jorgensen, et al., *J. Am. Chem. Soc.* (1996), v 118, pp 11225_11236; Jorgensen, W.L.; BOSS, Version 4.1; Yale University: New Haven, CT (1999)); OPLS (Jorgensen, et al., *J. Am. Chem. Soc.* (1988), v 110, pp 1657ff; Jorgensen, et al., *J. Am. Chem. Soc.* (1990), v 112, pp 4768ff); UNRES (United Residue Forcefield; Liwo, et al., *Protein Science* (1993), v 2, pp1697_1714; Liwo, et al., *Protein Science* (1993), v 2, pp1715_1731; Liwo, et al., *J. Comp. Chem.* (1997), v 18, pp849_873; Liwo, et al., *J. Comp. Chem.* (1997), v 18, pp874_884; Liwo, et al., *J. Comp. Chem.* (1998), v 19, pp259_276; Forcefield for Protein Structure Prediction (Liwo, et al., *Proc. Natl. Acad. Sci. USA* (1999), v 96, pp5482_5485); ECEPP/3 (Liwo et al., *J. Protein Chem* 1994 May 13(4):375_80); AMBER 1.1 force field (Weiner, et al., *J. Am. Chem. Soc.* v106, pp766_784); AMBER 3.0 force field (U.C. Singh et al., *Proc. Natl. Acad. Sci. USA.* 82:755_759); CHARMM and CHARMM22 (Brooks, et al., *J. Comp. Chem.* v4, pp 187_217); cvff3.0 (Dauber_Osguthorpe, et al., (1988) *Proteins: Structure, Function and Genetics*, v4, pp31_47); cff91 (Maple, et al., *J. Comp. Chem.* v15, 162_182); also, the DISCOVER (cvff and cff91) and AMBER forcefields are used in the INSIGHT molecular modeling package (Biosym/MSI, San Diego California) and HARMM is used in the QUANTA molecular modeling package (Biosym/MSI, San Diego California), all of which are expressly incorporated by reference. In fact, as is outlined below, these force field methods may be used to generate the secondary library directly; that is, no primary library is generated; rather, these methods can be used to generate a probability table from which the secondary library is directly generated, for example by using these force fields during an SCMF calculation.
- Once the singles and doubles energies are calculated and stored, the next step of the computational processing may occur. As outlined in U.S.S.N. 09/127,926 and PCT US98/07254, preferred embodiments utilize a Dead End Elimination (DEE) step, and preferably a Monte Carlo step.

WO 02/090300

PCT/US02/14358

PDA™, viewed broadly, has three components that may be varied to alter the output (e.g. the primary library): the scoring functions used in the process; the filtering technique, and the sampling technique.

In a preferred embodiment, the scoring functions may be altered. In a preferred embodiment, the scoring functions outlined above may be biased or weighted in a variety of ways. For example, a bias towards or away from a reference sequence or family of sequences can be done; for example, a bias towards wild-type or homolog residues may be used. Similarly, the entire protein or a fragment of it may be biased; for example, the active site may be biased towards wild-type residues, or domain residues towards a particular desired physical property can be done. Furthermore, a bias towards or against increased energy can be generated. Additional scoring function biases include, but are not limited to applying electrostatic potential gradients or hydrophobicity gradients, adding a substrate or binding partner to the calculation, or biasing towards a desired charge or hydrophobicity.

In addition, in an alternative embodiment, there are a variety of additional scoring functions that may be used. Additional scoring functions include, but are not limited to torsional potentials, or residue pair potentials, or residue entropy potentials. Such additional scoring functions can be used alone, or as functions for processing the library after it is scored initially. For example, a variety of functions derived from data on binding of peptides to MHC (Major Histocompatibility Complex) can be used to rescore a library in order to eliminate proteins containing sequences which can potentially bind to MHC, i.e. potentially immunogenic sequences.

In a preferred embodiment, a variety of filtering techniques can be done, including, but not limited to, DEE and its related counterparts. Additional filtering techniques include, but are not limited to branch-and-bound techniques for finding optimal sequences (Gordon and Majo, Structure Fold. Des. 7:1089-98, 1999), and exhaustive enumeration of sequences. It should be noted however, that some techniques may also be done without any filtering techniques; for example, sampling techniques can be used to find good sequences, in the absence of filtering.

As will be appreciated by those in the art, once an optimized sequence or set of sequences is generated, a variety of sequence space sampling methods can be done, either in addition to the preferred Monte Carlo methods, or instead of a Monte Carlo search. That is, once a sequence or set of sequences is generated, preferred methods utilize sampling techniques to allow the generation of additional, related sequences for testing.

These sampling methods can include the use of amino acid substitutions, insertions or deletions, or recombinations of one or more sequences. As outlined herein, a preferred embodiment utilizes a Monte Carlo search, which is a series of biased, systematic, or random jumps. However, there are other sampling techniques that can be used, including Boltzman sampling, genetic algorithm techniques and simulated annealing. In addition, for all the sampling techniques, the kinds of jumps

WO 02/090300

PCT/US02/14358

allowed can be altered (e.g. random jumps to random residues, biased jumps (to or away from wild-type, for example), jumps to biased residues (to or away from similar residues, for example, etc.). Jumps where multiple residue positions are coupled (two residues always change together, or never change together), jumps where whole sets of residues change to other sequences (e.g.,

5 recombination). Similarly, for all the sampling techniques, the acceptance criteria of whether a sampling jump is accepted can be altered, to allow broad searches at high temperature and narrow searches close to local optima at low temperatures. See Metropolis et al., J. Chem Phys v21, pp 1087, 1953, hereby expressly incorporated by reference.

10 In addition, it should be noted that the preferred methods of the invention result in a rank ordered list of sequences; that is, the sequences are ranked on the basis of some objective criteria. However, as outlined herein, it is possible to create a set of non-ordered sequences, for example by generating a probability table directly (for example using SCMF analysis or sequence alignment techniques) that lists sequences without ranking them. The sampling techniques outlined herein can be used in either

15 situation.

In a preferred embodiment, Boltzman sampling is done. As will be appreciated by those in the art, the temperature criteria for Boltzman sampling can be altered to allow broad searches at high temperature and narrow searches close to local optima at low temperatures (see e.g., Metropolis et

20 al., J. Chem. Phys. 21:1087, 1953).

In a preferred embodiment, the sampling technique utilizes genetic algorithms, e.g., such as those described by Holland (Adaptation in Natural and Artificial Systems, 1975, Ann Arbor, U. Michigan Press). Genetic algorithm analysis generally takes generated sequences and recombines them

25 computationally, similar to a nucleic acid recombination event, in a manner similar to "gene shuffling". Thus the "jumps" of genetic algorithm analysis generally are multiple position jumps. In addition, as outlined below, correlated multiple jumps may also be done. Such jumps can occur with different crossover positions and more than one recombination at a time, and can involve recombination of two or more sequences. Furthermore, deletions or insertions (random or biased) can be done. In

30 addition, as outlined below, genetic algorithm analysis may also be used after the secondary library has been generated.

In a preferred embodiment, the sampling technique utilizes simulated annealing, e.g., such as described by Kirkpatrick et al. [Science, 220:671-680 (1983)]. Simulated annealing alters the cutoff

35 for accepting good or bad jumps by altering the temperature. That is, the stringency of the cutoff is altered by altering the temperature. This allows broad searches at high temperature to new areas of sequence space, altering with narrow searches at low temperature to explore regions in detail.

WO 02/090300

PCT/US02/14358

In addition, there are computational methods that may be used as described in U.S.S.N.s 09/927,790, 60/352,103, and 60/351,937, all of which are expressly incorporated herein by reference.

Any protein design cycle can be used individually, in combination with other methods, or in
5 reiterations that combine methods.

In a preferred embodiment, the methods of the invention involve starting with a target protein and use experimental methods to generate a variant protein. That is, nucleic acid recombination techniques as are known to one of skill in the art are used to experimentally generate the variant proteins of the
10 present invention.

Thus, use of a nucleic acid recombination method or implementation of a protein design cycle, or a combination of nucleic acid recombination methods and computational processing results in the generation of a variant protein exhibiting altered cofactor specificity. By "variant protein" or "variable protein sequence" herein is meant a protein that differs from the scaffold protein or target protein in at
15 least one amino acid residue.

In a preferred embodiment, the cofactor specificity of the variant protein is altered compare to the target protein. Target proteins include but are not limited to thioredoxin reductase, glutathione reductase, and 2,5-diketo-D-gluconic acid reductase. Two specific amino acid regions have previously been reported for cofactor specificity (Carugo and Argos, *Proteins* (1997) 28, 10-28). The first region immediately follows the Gly-rich loop with the motif G-x-G-x-X₁-X₂, and is involved in pyridine nucleotide binding. Originally, it was believed that in proteins specific for NADPH, X₁ and X₂ are polar residues (Ser/Thr) and Ala, respectively, whereas for proteins specific for NADH, X₁ and X₂ are hydrophobic residues (Val/Ile) and Gly, respectively. The determination of additional sequences,
20 25 however, demonstrated significant sequence variability for X₁ and X₂, breaking this original rule for cofactor specificity.

The second region is reported as generally corresponding to the region from about amino acid 175 to amino acid 181 in *E.coli* thioredoxin reductase. In the NADH-dependent bacterial flavoprotein reductases Cp34 and AhpF (Reynolds et al., *Biochemistry* (2002) 41, 1990-2001), the second motif is reported as H-Q-F-x-x-x-Q and E-F-A-x-x-x-K, respectively. In a mutation study (Scrutton et al., *Nature* (1990) 343, 38-43; Mittl et al., *Protein Sci.* (1994) 3, 1504-1514), the NADPH specificity of *E.coli* GR was switched to NADH by mutation of the second motif to E-M-F-x-x-x-x-P (see picture
30 35 below).

In a preferred embodiment, a variant thioredoxin reductase is made in which the cofactor specificity is altered. Thioredoxin (TR) is a potent protein disulfide reductase found in most organisms that participates in many thiol-dependent cellular reductive processes. In addition to its ability to effect the

WO 02/090300

PCT/US02/14358

reduction of cellular proteins, it is recognized that thioredoxin reductase can act directly as an antioxidant (e.g., by preventing oxidation of an oxidizable substrate by scavenging reactive oxygen species) or can increase the oxidative stress in a cell by autooxidizing (e.g., generating superoxide radicals through autooxidation).

5

Thioredoxins are low molecular weight dithiol proteins that have the ability to reduce disulfides in typical organic compounds such as Ellman's reagent or disulfides as they exist naturally in a variety of proteins (Holmgren, A. (1981) *Trends in Biochemical Science*, 6, 26-39). Under normal physiological conditions, following the reduction of a disulfide bond, the oxidized thioredoxin is reduced by thioredoxin reductase, with the aid of NADPH as a cofactor. Thioredoxin of a species is typically reduced only by thioredoxin reductase of the same species.

10

The active site pocket of the thioredoxin reductases exhibits a highly conserved region across species, as shown in the amino acid alignment of Figure 1A. This region corresponds to the amino acid region between residues 156 and 181 of the *E. coli* thioredoxin reductase, or residues 220 and 245 of the *Arabidopsis* thioredoxin reductase. This highly conserved pocket is mostly responsible for the binding of the co-factor, NADPH. The trans-species variations in the amino acid sequence of thioredoxin reductase appear in the C- and N-termini regions, i.e., the region between residues 1-155 and 182-C-terminus of the *E. coli* thioredoxin reductase, or residues 1-219 and 246-C-terminus of the *Arabidopsis* thioredoxin reductase.

15

20

The target proteins used to generate the variant thioredoxin reductases of the present invention may be obtained from any organism including, but not limited to, *E. coli*, *Bacillus subtilis*, *Mycobacterium leprae*, *Sarcopharyngomyces*, *Neurospora crassa*, *Arabidopsis*, *Homo sapiens*, *Methanosarcina acetivorans* str. C2A, *Ureaplasma parvum*, *Mycoplasma pulmonis*, *Rickettsia conorii*, *Spironucleus barkhanus*, *Listeria innocua*, *Fusobacterium nucleatum*, *Methanococcus jannaschii*, *Mycoplasma genitalium*, *Haemophilus influenzae*, *Vibrio cholera*, *Listeria monocytogenes*, *Helicobacter pylori*, *Methanopyrus kandleri* AV19, *Schizosaccharomyces pombe*, *Chlamydomonas pneumoniae*, *Streptococcus pyogenes*, *Plasmodium falciparum*, *Mycobacterium tuberculosis*, *Mycoplasma genitalium*, *Borrelia burgdorferi*, *Ralstonia solanacearum*, *Sinorhizobium meliloti*, *Caulobacter crescentus* CB15J, *Encephalitozoon cuniculi*, *Staphylococcus aureus*, *Clostridium perfringens*, *Halobacterium* sp. NRC-1, *Sulfolobus solfataricus*, *Rickettsia prowazekii*, *Mesorhizobium loti*, *Mus musculus*, *Thermoplasma acidophilum*, *Sulfolobus tokodaii*, *Chlamydomonas pneumoniae*, *Mycoplasma pulmonis*, *Campylobacter jejuni*, *Chlamydia trachomatis*, *Aeropyrum pernix*, *Neisseria meningitidis*, *Pyrococcus horikoshii*, *Pyrococcus abyssi*, *Thermoplasma volcanium*, *Pyrococcus furiosus*, *Archaeoglobus fulgidus*, *Yersinia pestis*, *Bacillus halodurans*, *Ureaplasma urealyticum*, *Methanothermobacter thermotrophicus*, *Pyrobaculum aerophilum*, *Chlamydia muridarum*, *Treponema pallidum*, *Streptomyces coelicolor*, *Brucella melitensis*, *Agrobacterium tumefaciens*, *Drosophila melanogaster*, *Streptococcus pneumoniae*, *Clostridium acetobutylicum*, *Xylella fastidiosa*,

25

30

35

WO 02/090300

PCT/US02/14358

Lactococcus lactis, *Thermotoga maritima*, *Pseudomonas aeruginosa*, *Salmonella enterica*, *Nostoc* sp, *Deinococcus radiodurans*, *Penicillium chrysogenum*, *Salmonella typhimurium*, *Lactobacillus elbrueckii*, *Clostridium sticklandii*, *Clostridium litorale*, *Clostridium acetobutylicum*, *Thermoplasma volcanium*, *Rattus norvegicus*, *Coccidioides immitis*, *Bos Taurus*, *Mycobacterium smegmatis*,
 5 *Synechocystis* sp, *Plasmodium falciparum*, *Carboxydotherrmus hydrogenofomans*, *Sus scrofa*
Triticum aestivum.

In a preferred embodiment, the target proteins used to generate the variant thioredoxin reductases are selected from *E. coli*, *Bacillus subtilis*, *Mycobacterium leprae*, *Saccharomyces*, *Neurospora*
 10 *crassa*, *Arabidopsis*, *Homo sapiens*, barley TR found in US6380372, entitled Barley gene for
 Thioredoxin and NADP-thioredoxin reductase, issued 20020430; rice TR found in WO0198509 as
 amino acid sequence of SEQ ID NO:27 therein and its nucleotide sequence as sequence of SEQ ID
 NO:25 therein, the heat stable TRs from *Archaeoglobus fulgidus* (gil2649006) (trxB) which is the
 protein sequence SEQ ID NO:7 in WO0198509, and the protein sequence of TR from *Methanococcus*
 15 *jannaschii* (gil 1592167) (trxB), which is SEQ ID NO:6 in WO0198509.

In a preferred embodiment, the catalytic efficiency of the variant TR proteins is improved for the
 cofactor NADPH. Preferably, the catalytic efficiency of variant TRs is improved by at least about 5%
 as compared to wild-type for NADPH. More preferably, the catalytic efficiency of variant TRs is
 20 improved by at least about 15% as compared to wild-type for NADPH. More preferably the catalytic
 efficiency of variant TRs is improved by at least about 25% as compared to wild-type for NADPH.
 More preferably, the catalytic efficiency of variant TRs is improved by at least about 50% as
 compared to wild-type for NADPH. More preferably, the catalytic efficiency of variant TRs is improved
 by at least about 100% as compared to wild-type for NADPH. More preferably, the catalytic efficiency
 25 of variant TRs is improved by at least about 300% as compared to wild-type for NADPH. More
 preferably, the catalytic efficiency of variant TRs is improved by at least 500% as compared to wild-
 type for NADPH.

In a preferred embodiment, the catalytic efficiency of the variant TR proteins is improved for the
 cofactor NADH. Preferably, the catalytic efficiency of variant TRs is improved by at least about 5% as
 30 compared to wild-type for NADH. More preferably the catalytic efficiency of variant TRs is improved by
 at least about 15% as compared to wild-type for NADH. More preferably, the catalytic efficiency of
 variant TRs is improved by at least about 25% as compared to wild-type for NADH. More preferably,
 the catalytic efficiency of variant TRs is improved by at least about 50% as compared to wild-type for
 35 NADH. More preferably, the catalytic efficiency of variant TRs is improved by at least about 100% as
 compared to wild-type for NADH. More preferably, the catalytic efficiency of variant TRs is
 improved by at least about 300% as compared to wild-type for NADH. More preferably, the catalytic
 efficiency of variant TRs is improved by at least about 500% as compared to wild-type for NADH.
 More preferably, the catalytic efficiency of variant TRs is improved by at least about 1000% as

WO 02/090300

PCT/US02/14358

compared to wild-type for NADH. More preferably, the catalytic efficiency of variant TRs is improved by at least about 1300% as compared to wild-type for NADH. More preferably, the catalytic efficiency of variant TRs is improved by at least about 3000% as compared to wild-type for NADH.

- 5 In a preferred embodiment, the cofactor specificity of the variant thioredoxin reductase is altered such that there is an increased activity using NADH. Preferably, variant thioredoxin reductases (TR) will have at least 50% of native NADPH dependent activity using NADH. More preferably, variant thioredoxin reductases (TR) will have at least 75% of native NADPH dependent activity using NADH. More preferably, variant thioredoxin reductases (TR) will have at least 85% of native NADPH dependent activity using NADH. More preferably, variant thioredoxin reductases (TR) will have at least 95% of native NADPH dependent activity using NADH. More preferably, variant thioredoxin reductases (TR) will have at least 100% of native NADPH dependent activity using NADH.

- 15 In a preferred embodiment, the cofactor specificity of the variant thioredoxin reductase is altered such that there is a cofactor switch from NADPH to NADH. In other words, these variants will have an increase in NADH-dependent activity and a substantially simultaneous decrease in NADPH dependent activity. Preferably, variant thioredoxin reductase (TRs) will have at least 50% of native NADPH dependent activity using NADH. More preferably, variant thioredoxin reductase will have at least 75% of native NADPH dependent activity using NADH. More preferably, variant thioredoxin reductase will have at least 85% of native NADPH dependent activity using NADH. More preferably, variant thioredoxin reductase will have at least 95% of native NADPH dependent activity using NADH. More preferably, variant thioredoxin reductase will have at least 100% of native NADPH dependent activity using NADH.

- 25 Preferably, variant thioredoxin reductases (TRs) will have less than about 0.5% of native NADPH dependent activity. More preferably, TRs will have less than about 5% of native NADPH dependent activity. More preferably, TRs will have less than about 20% of native NADPH dependent activity. More preferably, TRs will have less than about 25% of native NADPH dependent activity. More preferably, TRs will have less than about 30% of native NADPH dependent activity. More preferably, TRs will have less than about 50% of native NADPH dependent activity. More preferably, TRs will have less than about 75% of native NADPH dependent activity. More preferably, TRs will have less than about 95% of native NADPH dependent activity.

- 35 In another embodiment, the catalytic efficiency of the variant TR proteins is improved for both co-factors, NADH and NADPH, together. Preferably, the catalytic efficiency of the TR variants is improved by at least about 5% as compared to wild-type for either of the two co-factors. More preferably, the catalytic efficiency of the TR variants is improved by at least about 50% as compared to wild-type for either of the two co-factors. More preferably, the catalytic efficiency of the TR variants is improved by at least about 100% as compared to wild-type for either of the two co-factors. More

WO 02/090300

PCT/US02/14358

preferably, the catalytic efficiency of the TR variants is improved by at least about 300% as compared to wild-type for either of the two co-factors. . More preferably, the catalytic efficiency of the TR variants is improved by at least about 1000% as compared to wild-type for either of the two co-factors. More preferably, the catalytic efficiency of the TR variants is improved by at least about 2000% as compared to wild-type for either of the two co-factors.

In a preferred embodiment, the NADPH binding affinity of the variant thioredoxin reductases (TRs) may be unaffected, reduced, or enhanced. For example, in some embodiments, variant TRs show greater than 100 times more affinity for NADPH than for NADH, while in other embodiments, variant TRs show greater than 50 times more affinity for NADPH than for NADH, or variant TRs may show greater than 25 times more affinity for NADPH than for NADH.

In a preferred embodiment, the ability of the variant TR protein to reduce its cognate thioredoxin is not substantially affected.

In a preferred embodiment, the substrate specificity of the variant TR protein is altered such that the TR protein may act on a thioredoxin protein from another species

In some embodiments, potential glycosylation sites added by protein design cycle may be removed without affecting activity by using a second protein design cycle.

In a preferred embodiment, the variant TR proteins have from 1 to 3 amino acid substitutions in amino acid regions involved in cofactor specificity as compared to the wild-type TR proteins. In other embodiments, the variant TR proteins have additional amino acid substitutions at other positions.

Thus, variant TR proteins may have at least about 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40 different residues in other positions. As will be appreciated by those of skill in the art, the number of additional positions that may have amino acid substitutions will depend on the wild-type TR protein used to generate the variants. Thus, in some instances, up to 50 different positions may have amino acid substitutions.

In a preferred embodiment, the variant TR protein comprise amino acid substitutions are selected from positions A4, A5 and A6, corresponding to positions 190, 191, and 195 in the *Arabidopsis* NTR protein (Genbank accession no. Q39242), positions 156, 157, and 175 in the *E. coli* TR protein (Genbank accession no P09625), positions 155, 156, and 174 in the *Bacillus subtilis* TR protein (Genbank accession no P80880), positions 163, 164, and 182 in the *Mycobacterium leprae* TR protein (Genbank accession no P46843), residue 164, 165, and 183 in the *Saccharomyces* TR protein (Genbank accession no P29509 and P38816), positions 163, 164, and 182 in the *Neurospora crassa* TR protein (Genbank accession no P51978), residue 170, 171, 189 in the *Arabidopsis* TR protein

WO 02/090300

PCT/US02/14358

(Genbank accession no Q39243) and residue 217, 218 and 249 in the Human TR protein (Genbank accession no Q16881).

5 In a preferred embodiment, the variant TR proteins comprise amino acid substitutions selected from the group of substitutions consisting of RA4W, RA5L, R A5M, R A5I, R A5F, R A5V, R A5Y, RA5A, RA5S, RA5C, RA5T, RA6T, R A6S, R A6Q, R A6G, and R A6N, RA6D, RA6M, and RA6E

10 In a preferred embodiment, the variant TR protein comprises the amino acid substitutions RA4W and RA6T.

15 In a preferred embodiment, the variant TR protein comprises the amino acid substitutions RA4W, RA5L, and RA6S.

20 In a preferred embodiment, the variant TR protein comprises the amino acid substitutions RA5Y and RA6N.

25 In a preferred embodiment, the variant TR protein comprises the amino acid substitutions RA4W, RA5F, and RA6Q.

30 In a preferred embodiment, the variant TR protein comprises the amino acid substitutions RA4W, RA5L, and RA6T.

35 In a preferred embodiment, the variant TR protein comprises the amino acid substitutions RA4W and RA6S.

40 In a preferred embodiment, the variant TR protein comprises the amino acid substitutions RA5Y and RA6N.

45 In a preferred embodiment, the variant TR protein comprises the amino acid substitutions RA5F and RA6N.

50 In a preferred embodiment, the variant TR protein comprises the amino acid substitutions RA4W and RA6T.

55 In a preferred embodiment, the variant TR protein comprises the amino acid substitutions RA4W, RA5L and RA6S.

60 In a preferred embodiment, the variant TR protein comprises the amino acid substitutions RA4W, RA5M, and RA6S.

WO 02/090300

PCT/US02/14358

In a preferred embodiment, the variant TR protein comprises the amino acid substitutions RA4W, RA5I, and RA6S.

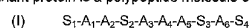
In a preferred embodiment, the variant TR protein comprises the amino acid substitutions RA4W, RA5F, and RA6Q.

In a preferred embodiment, the variant TR protein comprises the amino acid substitutions RA4W, and RA5V.

In a preferred embodiment, the variant TR protein comprises the amino acid substitutions RA4W, RA5M, and RA6G.

In a preferred embodiment, the variant TR protein comprises the amino acid substitutions RA4W, RA5V, and RA6G.

In a preferred embodiment, variant protein is a polypeptide molecule of Formula I.



where

- a) S_1 comprises a polypeptide sequence selected from the group consisting of SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, and SEQ ID NO:7, or a sequence having substantial similarity thereto;
- b) S_2 comprises a polypeptide sequence selected from the group consisting of SEQ ID NO:8, SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, and SEQ ID NO:14, or a sequence having substantial similarity thereto;
- c) S_3 comprises a polypeptide sequence selected from the group consisting of SEQ ID NO:15, SEQ ID NO:16, SEQ ID NO:17, SEQ ID NO:18, SEQ ID NO:19, SEQ ID NO:20, and SEQ ID NO:21, or a sequence having substantial similarity thereto;
- d) S_4 comprises a polypeptide sequence selected from the group consisting of SEQ ID NO:22, SEQ ID NO:23, SEQ ID NO:24, SEQ ID NO:25, SEQ ID NO:26, SEQ ID NO:27, and SEQ ID NO:28, or a sequence having substantial similarity thereto;
- e) A_1 is an amino acid moiety selected from the group consisting of serine, valine, glycine, alanine, leucine, isoleucine, methionine, phenylalanine, and tryptophan;
- f) A_2 is an amino acid moiety selected from the group consisting of alanine, glycine, valine, leucine, isoleucine, methionine, phenylalanine, and tryptophan;
- g) A_3 is an amino acid moiety selected from the group consisting of histidine, aspartic acid, glutamic acid, arginine, leucine, serine, threonine, cysteine, asparagine, glutamine, and tyrosine;
- h) A_4 is an amino acid moiety selected from the group consisting of arginine, alanine, glycine, valine, leucine, isoleucine, methionine, phenylalanine, and tryptophan;

WO 02/090300

PCT/US02/14358

- i) A_8 is an amino acid moiety selected from the group consisting of arginine, asparagine, glutamine, aspartic acid, glutamic acid, cysteine, serine, threonine, and lysine;
- j) A_9 is an amino acid moiety selected from the group consisting of arginine, glutamic acid, asparagine, glutamine, aspartic acid, cysteine, serine, threonine, and lysine;

provided that at least

- A_1 is not serine;
- A_2 is not alanine;
- A_3 is not histidine;
- A_4 is not arginine;
- A_5 is not arginine; or
- A_6 is not arginine.

In Formula I, above, the sequence $A_1-A_2-S_2-A_3-A_4-A_5-S_3-A_6$ corresponds to a highly conserved pocket in the sequence of thioredoxin reductase proteins obtained from various species. A_1 corresponds to residue 156 in the *E. coli* thioredoxin reductase sequence, residue 155 in the *Bacillus subtilis* thioredoxin reductase sequence, residue 163 in the *Mycobacterium leprae* thioredoxin reductase sequence, residue 164 in the *Sarccharomyces* thioredoxin reductase sequence, residue 163 in the *Neurospora crassa* thioredoxin reductase sequence, residue 170 in the *Arabidopsis* thioredoxin reductase sequence, and residue 217 in the Human thioredoxin reductase sequence. In the wild-type protein, this residue is threonine for *E. coli* and human, and serine for the other listed species.

A_2 corresponds to residue 157 in the *E. coli* thioredoxin reductase sequence, residue 156 in the *Bacillus subtilis* thioredoxin reductase sequence, residue 164 in the *Mycobacterium leprae* thioredoxin reductase sequence, residue 165 in the *Sarccharomyces* thioredoxin reductase sequence, residue 164 in the *Neurospora crassa* thioredoxin reductase sequence, residue 171 in the *Arabidopsis* thioredoxin reductase sequence, residue 218 in the Human thioredoxin reductase sequence. In the wild-type protein, this residue is valine for human and alanine for all the other listed species.

A_3 corresponds to residue 175 in the *E. coli* thioredoxin reductase sequence, residue 174 in the *Bacillus subtilis* thioredoxin reductase sequence, residue 182 in the *Mycobacterium leprae* thioredoxin reductase sequence, residue 183 in the *Sarccharomyces* thioredoxin reductase sequence, residue 182 in the *Neurospora crassa* thioredoxin reductase sequence, residue 189 in the *Arabidopsis* thioredoxin reductase sequence, residue 249 in the Human thioredoxin reductase sequence. In the wild-type protein, this residue is arginine for human, valine for *Sarccharomyces* and *Neurospora crassa*, and histidine for all the other listed species.

WO 02/090300

PCT/US02/14358

A₄ corresponds to residue 176 in the *E. coli* thioredoxin reductase sequence, residue 175 in the *Bacillus subtilis* thioredoxin reductase sequence, residue 183 in the *Mycobacterium leprae* thioredoxin reductase sequence, residue 184 in the *Sarcocharomyces* thioredoxin reductase sequence, residue 183 in the *Neurospora crassa* thioredoxin reductase sequence, residue 190 in the *Arabidopsis* thioredoxin reductase sequence, residue 250 in the Human thioredoxin reductase sequence. In the wild-type protein, this residue is glutamine for human and arginine for all the other listed species.

A₅ corresponds to residue 177 in the *E. coli* thioredoxin reductase sequence, residue 176 in the *Bacillus subtilis* thioredoxin reductase sequence, residue 184 in the *Mycobacterium leprae* thioredoxin reductase sequence, residue 185 in the *Sarcocharomyces* thioredoxin reductase sequence, residue 184 in the *Neurospora crassa* thioredoxin reductase sequence, residue 191 in the *Arabidopsis* thioredoxin reductase sequence, residue 251 in the Human thioredoxin reductase sequence. In the wild-type protein, this residue is lysine for *Sarcocharomyces* and *Neurospora crassa*, phenylalanine for human, and arginine for all the other listed species.

A₆ corresponds to residue 181 in the *E. coli* thioredoxin reductase sequence, residue 180 in the *Bacillus subtilis* thioredoxin reductase sequence, residue 188 in the *Mycobacterium leprae* thioredoxin reductase sequence, residue 189 in the *Sarcocharomyces* thioredoxin reductase sequence, residue 188 in the *Neurospora crassa* thioredoxin reductase sequence, residue 195 in the *Arabidopsis* thioredoxin reductase sequence, residue 255 in the Human thioredoxin reductase sequence. In the wild-type protein, this residue is lysine for human and arginine for all the other listed species.

It has been observed that among the species mentioned above, the portion of the amino acid sequence corresponding to S₂ and S₃ are also highly conserved. S₂ comprises a polypeptide sequence selected from the group consisting of SEQ ID NO:8, SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, and SEQ ID NO:14. S₃ comprises a polypeptide sequence selected from the group consisting of SEQ ID NO:15, SEQ ID NO:16, SEQ ID NO:17, SEQ ID NO:18, SEQ ID NO:19, SEQ ID NO:20, and SEQ ID NO:21 (Figure 2).

Therefore, embodiments of the invention relate to a polypeptide of Formula I, where S₁ consists of a polypeptide sequence having the sequence selected from the group consisting of SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, and SEQ ID NO:7.

In certain embodiments, S₂ consists of a polypeptide sequence selected from the group consisting of SEQ ID NO:8, SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, and SEQ ID NO:14, whereas S₃ consists of a polypeptide sequence selected from the group consisting of SEQ ID NO:15, SEQ ID NO:16, SEQ ID NO:17, SEQ ID NO:18, SEQ ID NO:19, SEQ ID NO:20, and

WO 02/090300

PCT/US02/14358

SEQ ID NO:21. Other embodiments of the invention relate to S_4 consisting of a polypeptide sequence having the sequence selected from the group consisting of SEQ ID NO:22, SEQ ID NO:23, SEQ ID NO:24, SEQ ID NO:25, SEQ ID NO:26, SEQ ID NO:27, and SEQ ID NO:28.

- 5 In one embodiment, in the polypeptide of Formula I, S_1 is the polypeptide sequence set forth in SEQ ID NO:1, S_2 is the polypeptide sequence set forth in SEQ ID NO:8, S_3 is the polypeptide sequence set forth in SEQ ID NO:15, and S_4 is the polypeptide sequence set forth in SEQ ID NO:22. This corresponds to a thioredoxin reductase protein, or a mutant thereof, obtained from *E. coli*.
- 10 In another embodiment, in the polypeptide of Formula I, S_1 is the polypeptide sequence set forth in SEQ ID NO:2, S_2 is the polypeptide sequence set forth in SEQ ID NO:9, S_3 is the polypeptide sequence set forth in SEQ ID NO:16, and S_4 is the polypeptide sequence set forth in SEQ ID NO:23. This corresponds to a thioredoxin reductase protein, or a mutant thereof, obtained from *Bacillus subtilis*.
- 15 In yet another embodiment, in the polypeptide of Formula I, S_1 is the polypeptide sequence set forth in SEQ ID NO:3, S_2 is the polypeptide sequence set forth in SEQ ID NO:10, S_3 is the polypeptide sequence set forth in SEQ ID NO:17, and S_4 is the polypeptide sequence set forth in SEQ ID NO:24. This corresponds to a thioredoxin reductase protein, or a mutant thereof, obtained from
- 20 *Mycobacterium leprae*.

- Another embodiment of the invention relates to a polypeptide of Formula I, in which S_1 is the polypeptide sequence set forth in SEQ ID NO:4, S_2 is the polypeptide sequence set forth in SEQ ID NO:11, S_3 is the polypeptide sequence set forth in SEQ ID NO:18, and S_4 is the polypeptide sequence set forth in SEQ ID NO:25. This corresponds to a thioredoxin reductase protein, or a mutant thereof, obtained from *Sarcomyces*.
- 25

- In another embodiment, in the polypeptide of Formula I, S_1 is the polypeptide sequence set forth in SEQ ID NO:5, S_2 is the polypeptide sequence set forth in SEQ ID NO:12, S_3 is the polypeptide sequence set forth in SEQ ID NO:19, and S_4 is the polypeptide sequence set forth in SEQ ID NO:26. This corresponds to a thioredoxin reductase protein, or a mutant thereof, obtained from *Neurospora crassa*.
- 30

- In one embodiment, in the polypeptide of Formula I, S_1 is the polypeptide sequence set forth in SEQ ID NO:6, S_2 is the polypeptide sequence set forth in SEQ ID NO:13, S_3 is the polypeptide sequence set forth in SEQ ID NO:20, and S_4 is the polypeptide sequence set forth in SEQ ID NO:27. This corresponds to a thioredoxin reductase protein, or a mutant thereof, obtained from *Arabidopsis*.
- 35

WO 02/090300

PCT/US02/14358

The invention also relates to another polypeptide of Formula I, in which S₁ is the polypeptide sequence set forth in SEQ ID NO:7, S₂ is the polypeptide sequence set forth in SEQ ID NO:14, S₃ is the polypeptide sequence set forth in SEQ ID NO:21, and S₄ is the polypeptide sequence set forth in SEQ ID NO:28. This corresponds to a thioredoxin reductase protein, or a mutant thereof, obtained from Human.

The invention encompasses certain mutants of the naturally occurring thioredoxin reductase proteins. These mutants include those in which A₁ is an amino acid moiety selected from the group consisting of valine, alanine, and leucine; A₂ is an amino acid moiety selected from the group consisting of glycine, valine, and leucine; A₃ is an amino acid moiety selected from the group consisting of aspartic acid, glutamic acid, asparagine, and glutamine; A₄ is an amino acid moiety selected from the group consisting of alanine, glycine, valine, leucine, isoleucine, and methionine; A₅ is an amino acid moiety selected from the group consisting of asparagine, glutamine, aspartic acid, and glutamic acid; A₆ is an amino acid moiety selected from the group consisting of glutamic acid, glutamine, aspartic acid, and asparagine.

It is understood that a polypeptide of the present invention may have one or more than one of the above mutations.

In certain embodiments A₁ is valine, while in others A₂ is glycine, and in others A₃ is aspartic acid; and in others A₄ is alanine, while in others A₅ is asparagine, and in others A₆ is glutamic acid. In some embodiments, two or more of these particular amino acid residues exist at the specified position.

In a preferred embodiment the variant proteins of the present invention may be fused to a second protein. For example, a fusion protein comprising the polypeptide of Formula I and a second polypeptide may be made. The second polypeptide may be a wild-type TR protein, wild-type thioredoxin, or a variant designed by a protein design cycle. Alternatively, a fusion protein comprising a variant protein generated by a protein design cycle and a second polypeptide may be fused. The second polypeptide may be a wild-type TR protein, wild-type thioredoxin or the polypeptide of Formula I. Such fusion may be through a linker.

By "linker", "linker sequence", "spacer", "tethering sequence" or grammatical equivalents thereof, herein is meant a molecule or group of molecules (such as a monomer or polymer) that connects two molecules and often serves to place the two molecules in a preferred configuration. In one aspect of this embodiment, the linker is a peptide bond. Choosing a suitable linker for a specific case where two polypeptide chains are to be connected depends on various parameters, e.g., the nature of the two polypeptide chains (e.g., whether they naturally form a dimer or not), the distance between the N- and the C-termini to be connected if known from three-dimensional structure determination, and/or the stability of the linker towards proteolysis and oxidation. Furthermore, the linker may contain amino

WO 02/090300

PCT/US02/14358

acid residues that provide flexibility. Thus, the linker peptide may predominantly include the following amino acid residues: Gly, Ser, Ala, or Thr.

5 The linker peptide should have a length that is adequate to link two monomers in such a way that they assume the correct conformation relative to one another so that they retain the desired activity as antagonists of a given receptor. Suitable lengths for this purpose includes at least one and not more than 30 amino acid residues. Preferably, the linker is from about 1 to 30 amino acids in length, with linkers of 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18 19 and 20 amino acids in length being preferred. See also WO 01/25277, incorporated herein by reference in its entirety.

10 In addition, the amino acid residues selected for inclusion in the linker peptide should exhibit properties that do not interfere significantly with the activity of the polypeptide. Thus, the linker peptide on the whole should not exhibit a charge that would be inconsistent with the activity of the polypeptide, or interfere with internal folding, or form bonds or other interactions with amino acid residues in one or more of the monomers that would seriously impede the binding of receptor monomer domains.

15 Useful linkers include glycine-serine polymers (including, for example, (GS)_n, (GSGGS)_n (GGGGS)_n and (GGGS)_n, where n is an integer of at least one), glycine-alanine polymers, alanine-serine polymers, and other flexible linkers such as the tether for the shaker potassium channel, and a large variety of other flexible linkers, as will be appreciated by those in the art. Glycine-serine polymers are preferred since both of these amino acids are relatively unstructured, and therefore may be able to serve as a neutral tether between components. Secondly, serine is hydrophilic and therefore able to solubilize what could be a globular glycine chain. Third, similar chains have been shown to be effective in joining subunits of recombinant proteins such as single chain antibodies.

20 Suitable linkers may also be identified by screening databases of known three-dimensional structures for naturally occurring motifs that can bridge the gap between two polypeptide chains. Another way of obtaining a suitable linker is by optimizing a simple linker, e.g., (Gly₂Ser)_n, through random mutagenesis.

25 In a preferred embodiment, the linker may comprise a polypeptide sequence having between about 5 and about 50 amino acids, or between about 10 and about 40 amino acids, or between about 15 and about 25 amino acids. In a preferred embodiment, the linker is about 22 amino acids.

30 In a preferred embodiment, the variant proteins of the present invention may be fused to a third polypeptide, and again, such fusion may be through a linker. The linker between the fusion polypeptide, which includes the polypeptide of Formula I, and the third polypeptide may have a molecular weight between about 5 and about 100 kDa, or a molecular weight between about 20 and

WO 02/090300

PCT/US02/14358

about 70 kDa, or even a molecular weight between about 25 and about 45 kDa. In a preferred embodiment, the linker has a molecular weight of between about 30 to about 40 kDa. In certain embodiments, this linker comprises amino acid residues that are negatively charged, such as glutamate and aspartate.

5

In certain embodiments, the third polypeptide is oleosin.

Thus, one embodiment of the present invention relates to a polypeptide of Formula I, which is fused to a second polypeptide at its C-terminus, perhaps through a linker, and is also fused to a third polypeptide at its N-terminus, again perhaps through another linker. Another embodiment of the invention relates to a series of fused polypeptides of Formula II

10

(II) oleosin-linker 1-thioredoxin reductase-linker 2-thioredoxin

15 where "linker 1" refers to the linker between the polypeptide of Formula I and the third polypeptide, set forth above, and "linker 2" refers to the linker between the polypeptide of Formula I and the second polypeptide, set forth above. Likewise, some embodiments of the invention can include any other fusion protein comprising the polypeptide of Formula I, whether it is fused to another protein at its N-terminus, its C-terminus, or both. Specifically, the invention contemplates modifications of Formula II or any other fusion of two polypeptides to the polypeptide of Formula I in which the components occur in any order.

20

In a preferred embodiment, the binding affinities of variant TR proteins for NADPH and NADH are determined. Suitable assays include, but are not limited to, e.g., quantitative comparisons comparing kinetic and equilibrium binding constants. The kinetic association rate (K_{on}) and dissociation rate (K_{off}), and the equilibrium binding constants (K_d) can be determined using surface plasmon resonance on a BIAcore instrument following the standard procedure in the literature [Pearce et al., Biochemistry 38:81-89 (1999)].

25

30 In a preferred embodiment, the antigenic profile in the host animal of the variant TR protein is similar, and preferably identical, to the antigenic profile of the host TR that is, the variant TR protein does not significantly stimulate the host organism (e.g. the patient) to an immune response; that is, any immune response is not clinically relevant and there is no allergic response or neutralization of the protein by an antibody. That is, in a preferred embodiment, the variant TR protein does not contain additional or different epitopes from the TR. By "epitope" or "determinant" herein is meant a portion of a protein which will generate and/or bind an antibody. Thus, in most instances, no significant amounts of antibodies are generated to a variant TR protein. In general, this is accomplished by not significantly altering surface residues, or by adding any amino acid residues on the surface which can become glycosylated, as novel glycosylation can result in an immune response.

35

WO 02/090300

PCT/US02/14358

The variant TR proteins and nucleic acids of the invention are distinguishable from naturally occurring wild-type TR. By "naturally occurring" or "wild type" or grammatical equivalents, herein is meant an amino acid sequence or a nucleotide sequence that is found in nature and includes allelic variations; that is, an amino acid sequence or a nucleotide sequence that usually has not been intentionally modified. Accordingly, by "non-naturally occurring" or "synthetic" or "recombinant" or grammatical equivalents thereof, herein is meant an amino acid sequence or a nucleotide sequence that is not found in nature; that is, an amino acid sequence or a nucleotide sequence that usually has been intentionally modified. It is understood that once a recombinant nucleic acid is made and reintroduced into a host cell or organism, it will replicate non-recombinantly, i.e., using the in vivo cellular machinery of the host cell rather than in vitro manipulations, however, such nucleic acids, once produced recombinantly, although subsequently replicated non-recombinantly, are still considered recombinant for the purpose of the invention. Representative amino acid sequences of naturally occurring TR proteins are shown in Figure 21. It should be noted that unless otherwise stated, all positional numbering of variant TR proteins and variant TR proteins is based on these sequences. That is, as will be appreciated by those in the art, an alignment of TR proteins and variant TR proteins can be done using standard programs, as is outlined below, with the identification of "equivalent" positions between the two proteins.

Thus, in a preferred embodiment, the variant TR protein has an amino acid sequence that differs from a wild-type TR sequence (Figure 21) by at least 1-5% of the residues. That is, the variant TR proteins of the invention are less than about 97-99% identical to a wild-type TR amino acid sequence. Accordingly, a protein is a "variant TR protein" if the overall homology of the protein sequence to the amino acid sequence is preferably less than about 99%, more preferably less than about 98%, even more preferably less than about 97% and more preferably less than 95% of a wild-type TR protein. In some embodiments, the homology will be as low as about 75-80%. Stated differently, variant TR proteins have at least about 1 residue that differs from the wild-type TR sequence (i.e., Figure 21), with at least about 2, 3, 4, 5, up to 50 different residues. Preferably variant TR proteins have 1 to 3 different residues. More preferably, variant TR proteins have 3 to 5 different residues. Preferably variant TR proteins have 5 to 10 different residues. Preferably variant TR proteins have 10 to 15 different residues. Preferably variant TR proteins have 15 to 25 different residues. Preferably variant TR proteins have 25 to 35 different residues.

Homology in this context means sequence similarity or identity, with identity being preferred. As is known in the art, a number of different programs can be used to identify whether a protein (or nucleic acid as discussed below) has sequence identity or similarity to a known sequence. Sequence identity and/or similarity is determined using standard techniques known in the art, including, but not limited to, the local sequence identity algorithm of Smith & Waterman, Adv. Appl. Math., 2:482 (1981), by the sequence identity alignment algorithm of Needleman & Wunsch, J. Mol. Biol., 48:443 (1970), by the search for similarity method of Pearson & Lipman, Proc. Natl. Acad. Sci. U.S.A., 85:2444 (1988), by

WO 02/090300

PCT/US02/14358

computerized implementations of these algorithms (GAP, BESTFIT, FASTA, and TFASTA in the Wisconsin Genetics Software Package, Genetics Computer Group, 575 Science Drive, Madison, WI), the Best Fit sequence program described by Devereux et al., Nucl. Acid Res., 12:387-395 (1984), preferably using the default settings, or by inspection. Preferably, percent identity is calculated by

5 FastDB based upon the following parameters: mismatch penalty of 1; gap penalty of 1; gap size penalty of 0.33; and joining penalty of 30, "Current Methods in Sequence Comparison and Analysis," Macromolecule Sequencing and Synthesis, Selected Methods and Applications, pp 127-149 (1988), Alan R. Liss, Inc.

- 10 An example of a useful algorithm is PILEUP. PILEUP creates a multiple sequence alignment from a group of related sequences using progressive, pair wise alignments. It can also plot a tree showing the clustering relationships used to create the alignment. PILEUP uses a simplification of the progressive alignment method of Feng & Doolittle, J. Mol. Evol. 35:351-360 (1987); the method is similar to that described by Higgins & Sharp CABIOS 5:151-153 (1989). Useful PILEUP parameters
- 15 including a default gap weight of 3.00, a default gap length weight of 0.10, and weighted end gaps.

- Another example of a useful algorithm is the BLAST algorithm, described in: Altschul et al., J. Mol. Biol. 215, 403-410, (1990); Altschul et al., Nucleic Acids Res. 25:3389-3402 (1997); and Karlin et al., Proc. Natl. Acad. Sci. U.S.A. 90:5873-5787 (1993). A particularly useful BLAST program is the WU-
- 20 BLAST-2 program which was obtained from Altschul et al., Methods in Enzymology, 266:460-480 (1996); [http://blast.wustl.edu/blast/ README.html](http://blast.wustl.edu/blast/README.html). WU-BLAST-2 uses several search parameters, most of which are set to the default values. The adjustable parameters are set with the following values: overlap span =1, overlap fraction = 0.125, word threshold (T) = 11. The HSP S and HSP S2 parameters are dynamic values and are established by the program itself depending upon the
- 25 composition of the particular sequence and composition of the particular database against which the sequence of interest is being searched; however, the values may be adjusted to increase sensitivity. An additional useful algorithm is gapped BLAST as reported by Altschul et al., Nucl. Acids Res., 25:3389-3402. Gapped BLAST uses BLOSUM-62 substitution scores; threshold T parameter set to 9; the two-hit method to trigger ungapped extensions; charges gap lengths of k a cost of 10+k; X_0 set
- 30 to 16, and X_1 set to 40 for database search stage and to 67 for the output stage of the algorithms. Gapped alignments are triggered by a score corresponding to ~22 bits.

- A % amino acid sequence identity value is determined by the number of matching identical residues divided by the total number of residues of the "longer" sequence in the aligned region. The "longer"
- 35 sequence is the one having the most actual residues in the aligned region (gaps introduced by WU-Blast-2 to maximize the alignment score are ignored).

In a similar manner, "percent (%) nucleic acid sequence identity" with respect to the coding sequence of the polypeptides identified herein is defined as the percentage of nucleotide residues in a candidate

WO 02/090300

PCT/US02/14358

sequence that are identical with the nucleotide residues in the coding sequence of the cell cycle protein. A preferred method utilizes the BLASTN module of WU-BLAST-2 set to the default parameters, with overlap span and overlap fraction set to 1 and 0.125, respectively.

- 5 The alignment may include the introduction of gaps in the sequences to be aligned. In addition, for sequences which contain either more or fewer amino acids than a wild-type TR sequence (i.e., see Figure 2, Figure 16N), it is understood that in one embodiment, the percentage of sequence identity will be determined based on the number of identical amino acids in relation to the total number of amino acids. Thus, for example, sequence identity of sequences shorter than a wild-type TR protein
- 10 sequence (i.e., see Figure 2, Figure 16N), as discussed below, will be determined using the number of amino acids in the shorter sequence, in one embodiment. In percent identity calculations relative weight is not assigned to various manifestations of sequence variation, such as, insertions, deletions, substitutions, etc.
- 15 In one embodiment, only identities are scored positively (+1) and all forms of sequence variation including gaps are assigned a value of "0", which obviates the need for a weighted scale or parameters as described below for sequence similarity calculations. Percent sequence identity can be calculated, for example, by dividing the number of matching identical residues by the total number of residues of the "shorter" sequence in the aligned region and multiplying by 100. The "longer"
- 20 sequence is the one having the most actual residues in the aligned region.

Thus, the variant TR proteins of the present invention may be shorter or longer than the amino acid sequence of wild-type TR proteins (i.e., Figure 21). Thus, in a preferred embodiment, included within the definition of variant TR proteins are portions or fragments of the sequences depicted herein.

- 25 Fragments of variant TR proteins are considered variant TR proteins if a) they share at least one antigenic epitope; b) have at least the indicated homology; c) and preferably have variant TR biological activity as defined herein.

- 30 In a preferred embodiment, as is more fully outlined below, the variant TR proteins include further amino acid variations, as compared to a wild type TR, than those outlined herein. In addition, as outlined herein, any of the variations depicted herein may be combined in any way to form additional novel variant TR proteins.

- 35 In addition, variant TR proteins can be made that are longer than those depicted in the figures, for example, by the addition of epitope or purification tags, as outlined herein, the addition of other fusion sequences, etc. For example, the variant TR proteins of the invention may be fused to other therapeutic proteins or to other proteins such as Fc or serum albumin for pharmacokinetic purposes. See for example U.S. Patent No. 5,766,883 and 5,876,969, both of which are expressly incorporated by reference.

WO 02/090300

PCT/US02/14358

In a preferred embodiment, the variant TR proteins of the invention are human TR conformers. By "conformer" herein is meant a protein that has a protein backbone 3D structure that is virtually the same but has significant differences in the amino acid side chains. That is, the variant TR proteins of the invention define a conformer set, wherein all of the proteins of the set share a backbone structure and yet have sequences that differ by at least 1-3-5%. The three dimensional backbone structure of a variant TR protein thus substantially corresponds to the three dimensional backbone structure of human TR. "Backbone" in this context means the non-side chain atoms: the nitrogen, carbonyl carbon and oxygen, and the α -carbon, and the hydrogens attached to the nitrogen and α -carbon. To be considered a conformer, a protein must have backbone atoms that are no more than 2 angstroms from the human TR structure, with no more than 1.5 angstroms being preferred, and no more than 1 angstrom being particularly preferred. In general, these distances may be determined in two ways. In one embodiment, each potential conformer is crystallized and its three dimensional structure determined. Alternatively, as the former is quite tedious, the sequence of each potential conformer is run in the PDA program to determine whether it is a conformer.

In alternative embodiments, the variant TR proteins of the invention may be conformers of any of the TR proteins listed in Figure 21.

Variant TR proteins may also be identified as being encoded by variant TR nucleic acids. In the case of the nucleic acid, the overall homology of the nucleic acid sequence is commensurate with amino acid homology but takes into account the degeneracy in the genetic code and codon bias of different organisms. Accordingly, the nucleic acid sequence homology may be either lower or higher than that of the protein sequence, with lower homology being preferred.

In a preferred embodiment, a variant TR nucleic acid encodes a variant TR protein. As will be appreciated by those in the art, due to the degeneracy of the genetic code, an extremely large number of nucleic acids may be made, all of which encode the variant TR proteins of the present invention. Thus, having identified a particular amino acid sequence, those skilled in the art could make any number of different nucleic acids, by simply modifying the sequence of one or more codons in a way which does not change the amino acid sequence of the variant TR.

In one embodiment, the nucleic acid homology is determined through hybridization studies. Thus, for example, nucleic acids which hybridize under high stringency to the nucleic acid sequence shown in Figure 21 or its complement and encode a variant TR protein is considered a variant TR gene.

High stringency conditions are known in the art; see for example Maniatis et al., Molecular Cloning: A Laboratory Manual, 2d Edition, 1989, and Short Protocols in Molecular Biology, ed. Ausubel, et al., both of which are hereby incorporated by reference. Stringent conditions are sequence-dependent and will be different in different circumstances. Longer sequences hybridize specifically at higher

WO 02/090300

PCT/US02/14358

- temperatures. An extensive guide to the hybridization of nucleic acids is found in Tijssen, Techniques in Biochemistry and Molecular Biology—Hybridization with Nucleic Acid Probes, "Overview of principles of hybridization and the strategy of nucleic acid assays" (1993). Generally, stringent conditions are selected to be about 5-10°C lower than the thermal melting point (T_m) for the specific sequence at a defined ionic strength and pH. The T_m is the temperature (under defined ionic strength, pH and nucleic acid concentration) at which 50% of the probes complementary to the target hybridize to the target sequence at equilibrium (as the target sequences are present in excess, at T_m , 50% of the probes are occupied at equilibrium). Stringent conditions will be those in which the salt concentration is less than about 1.0 M sodium ion, typically about 0.01 to 1.0 M sodium ion concentration (or other salts) at pH 7.0 to 8.3 and the temperature is at least about 30°C for short probes (e.g. 10 to 50 nucleotides) and at least about 60°C for long probes (e.g. greater than 50 nucleotides). Stringent conditions may also be achieved with the addition of destabilizing agents such as formamide.
- 15 In another embodiment, less stringent hybridization conditions are used; for example, moderate or low stringency conditions may be used, as are known in the art; see Maniatis and Ausubel, *supra*, and Tijssen, *supra*.
- The variant TR proteins and nucleic acids of the present invention are recombinant. As used herein, "nucleic acid" may refer to either DNA or RNA, or molecules which contain both deoxy- and ribonucleotides. The nucleic acids include genomic DNA, cDNA and oligonucleotides including sense and anti-sense nucleic acids. Such nucleic acids may also contain modifications in the ribose-phosphate backbone to increase stability and half life of such molecules in physiological environments.
- 25 The nucleic acid may be double stranded, single stranded, or contain portions of both double stranded or single stranded sequence. As will be appreciated by those in the art, the depiction of a single strand ("Watson") also defines the sequence of the other strand ("Crick"); thus the sequence depicted in Figure 6 also includes the complement of the sequence. By the term "recombinant nucleic acid" herein is meant nucleic acid, originally formed *in vitro*, in general, by the manipulation of nucleic acid by endonucleases, in a form not normally found in nature. Thus an isolated variant TR nucleic acid, in a linear form, or an expression vector formed *in vitro* by ligating DNA molecules that are not normally joined, are both considered recombinant for the purposes of this invention. It is understood that once a recombinant nucleic acid is made and reintroduced into a host cell or organism, it will replicate non-recombinantly, i.e. using the *in vivo* cellular machinery of the host cell rather than *in vitro* manipulations; however, such nucleic acids, once produced recombinantly, although subsequently replicated non-recombinantly, are still considered recombinant for the purposes of the invention.

WO 02/090300

PCT/US02/14358

Similarly, a "recombinant protein" is a protein made using recombinant techniques, i.e. through the expression of a recombinant nucleic acid as depicted above. A recombinant protein is distinguished from naturally occurring protein by at least one or more characteristics. For example, the protein may be isolated or purified away from some or all of the proteins and compounds with which it is normally associated in its wild type host, and thus may be substantially pure. For example, an isolated protein is unaccompanied by at least some of the material with which it is normally associated in its natural state, preferably constituting at least about 0.5%, more preferably at least about 5% by weight of the total protein in a given sample. A substantially pure protein comprises at least about 75% by weight of the total protein, with at least about 80% being preferred, and at least about 90% being particularly preferred. The definition includes the production of a variant TR protein from one organism in a different organism or host cell. Alternatively, the protein may be made at a significantly higher concentration than is normally seen, through the use of a inducible promoter or high expression promoter, such that the protein is made at increased concentration levels. Furthermore, all of the variant TR proteins outlined herein are in a form not normally found in nature, as they contain amino acid substitutions, insertions and deletions, with substitutions being preferred, as discussed below.

Also included within the definition of variant TR proteins of the present invention are amino acid sequence variants of the variant TR sequences outlined herein and shown in the Figures. That is, the variant TR proteins may contain additional variable positions as compared to human TR. These variants fall into one or more of three classes: substitutional, insertional or deletional variants. These variants ordinarily are prepared by site specific mutagenesis of nucleotides in the DNA encoding a variant TR protein, using cassette or PCR mutagenesis or other techniques well known in the art, to produce DNA encoding the variant, and thereafter expressing the DNA in recombinant cell culture as outlined above. However, variant TR protein fragments having up to about 100-150 residues may be prepared by *in vitro* synthesis using established techniques. Amino acid sequence variants are characterized by the predetermined nature of the variation, a feature that sets them apart from naturally occurring allelic or interspecies variation of the variant TR protein amino acid sequence. The variants typically exhibit the same qualitative biological activity as the naturally occurring analogue; although variants can also be selected which have modified characteristics as will be more fully outlined below.

While the site or region for introducing an amino acid sequence variation is predetermined, the mutation per se need not be predetermined. For example, in order to optimize the performance of a mutation at a given site, random mutagenesis may be conducted at the target codon or region and the expressed variant TR proteins screened for the optimal combination of desired activity. Techniques for making substitution mutations at predetermined sites in DNA having a known sequence are well known, for example, M13 primer mutagenesis and PCR mutagenesis. Screening of the mutants is done using assays of variant TR protein activities.

WO 02/090300

PCT/US02/14358

Amino acid substitutions are typically of single residues; insertions usually will be on the order of from about 1 to 20 amino acids, although considerably larger insertions may be tolerated. Deletions range from about 1 to about 20 residues, although in some cases deletions may be much larger.

- 5 Substitutions, deletions, insertions or any combination thereof may be used to arrive at a final derivative. Generally these changes are done on a few amino acids to minimize the alteration of the molecule. However, larger changes may be tolerated in certain circumstances. When small alterations in the characteristics of the variant TR protein are desired, substitutions are generally made in accordance with the following chart:

10

Chart 1

<i>Original Residue</i>	<i>Exemplary Substitutions</i>
Ala	Ser
Arg	Lys
Asn	Gln, His
Asp	Glu
Cys	Ser, Ala
Gln	Asn
Glu	Asp
Gly	Pro
His	Asn, Gln
Ile	Leu, Val
Leu	Ile, Val
Lys	Arg, Gln, Glu
Met	Leu, Ile
Phe	Met, Leu, Tyr
Ser	Thr
Thr	Ser
Trp	Tyr
Tyr	Trp, Phe
Val	Ile, Leu

- 15 Substantial changes in function or immunological identity are made by selecting substitutions that are less conservative than those shown in Chart I. For example, substitutions may be made which more significantly affect: the structure of the polypeptide backbone in the area of the alteration, for example the alpha-helical or beta-sheet structure; the charge or hydrophobicity of the molecule at the target site; or the bulk of the side chain. The substitutions which in general are expected to produce the greatest changes in the polypeptide's properties are those in which (a) a hydrophilic residue, e.g. seryl or threonyl, is substituted for (or by) a hydrophobic residue, e.g. leucyl, isoleucyl, phenylalanyl, valyl or alanyl; (b) a cysteine or proline is substituted for (or by) any other residue; (c) a residue having an electropositive side chain, e.g. lysyl, arginyl, or histidyl, is substituted for (or by) an electronegative residue, e.g. glutamyl or aspartyl; or (d) a residue having a bulky side chain, e.g. phenylalanine, is substituted for (or by) one not having a side chain, e.g. glycine.

25

WO 02/090300

PCT/US02/14358

The variants typically exhibit the same qualitative biological activity and will elicit the same immune response as the original variant TR protein, although variants also are selected to modify the characteristics of the variant TR proteins as needed. Alternatively, the variant may be designed such that the biological activity of the variant TR protein is altered. For example, glycosylation sites may be altered or removed. Similarly, the biological function may be altered; for example, in some instances it may be desirable to have more or less potent TR activity.

The variant TR proteins and nucleic acids of the invention can be made in a number of ways. Individual nucleic acids and proteins can be made as known in the art and outlined below.

Alternatively, libraries of variant TR proteins can be made for testing.

In a preferred embodiment, sets or libraries of variant TR proteins are generated from a probability distribution table. As outlined herein, there are a variety of methods of generating a probability distribution table, including using PDA, sequence alignments, forcefield calculations such as SCMF calculations, etc. In addition, the probability distribution can be used to generate information entropy scores for each position, as a measure of the mutational frequency observed in the library.

In this embodiment, the frequency of each amino acid residue at each variable position in the list is identified. Frequencies can be thresholded, wherein any variant frequency lower than a cutoff is set to zero. This cutoff is preferably 1%, 2%, 5%, 10% or 20%, with 10% being particularly preferred. These frequencies are then built into the variant TR library. That is, as above, these variable positions are collected and all possible combinations are generated, but the amino acid residues that "fill" the library are utilized on a frequency basis. Thus, in a non-frequency based library, a variable position that has 5 possible residues will have 20% of the proteins comprising that variable position with the first possible residue, 20% with the second, etc. However, in a frequency based library, a variable position that has 5 possible residues with frequencies of 10%, 15%, 25%, 30% and 20%, respectively, will have 10% of the proteins comprising that variable position with the first possible residue, 15% of the proteins with the second residue, 25% with the third, etc. As will be appreciated by those in the art, the actual frequency may depend on the method used to actually generate the proteins; for example, exact frequencies may be possible when the proteins are synthesized. However, when the frequency-based primer system outlined below is used, the actual frequencies at each position will vary, as outlined below.

As will be appreciated by those in the art and outlined herein, probability distribution tables can be generated in a variety of ways. In addition to the methods outlined herein, self-consistent mean field (SCMF) methods can be used in the direct generation of probability tables. SCMF is a deterministic computational method that uses a mean field description of rotamer interactions to calculate energies. A probability table generated in this way can be used to create libraries as described herein. SCMF can be used in three ways: the frequencies of amino acids and rotamers for each amino acid are

WO 02/090300

PCT/US02/14358

listed at each position; the probabilities are determined directly from SCMF (see Delarue et al. Pac. Symp. Biocomput. 109-21 (1997), expressly incorporated by reference). In addition, highly variable positions and non-variable positions can be identified. Alternatively, another method is used to determine what sequence is jumped to during a search of sequence space; SCMF is used to obtain an accurate energy for that sequence; this energy is then used to rank it and create a rank-ordered list of sequences (similar to a Monte Carlo sequence list). A probability table showing the frequencies of amino acids at each position can then be calculated from this list (Koehl et al., J. Mol. Biol. 239:249 (1994); Koehl et al., Nat. Struct. Biol. 2:163 (1995); Koehl et al., Curr. Opin. Struct. Biol. 6:222 (1996); Koehl et al., J. Mol. Bio. 293:1183 (1999); Koehl et al., J. Mol. Biol. 293:1161 (1999); Lee J. Mol. Biol. 236:918 (1994); and Vasquez Biopolymers 36:53-70 (1995); all of which are expressly incorporated by reference. Similar methods include, but are not limited to, OPLS-AA (Jorgensen, et al., J. Am. Chem. Soc. (1996), v 118, pp 11225-11236; Jorgensen, W.L.; BOSS, Version 4.1; Yale University: New Haven, CT (1999)); OPLS (Jorgensen, et al., J. Am. Chem. Soc. (1988), v 110, pp 1657ff; Jorgensen, et al., J. Am. Chem. Soc. (1990), v 112, pp 4768ff); UNRES (United Residue Forcefield; Liwo, et al., Protein Science (1993), v 2, pp1697-1714; Liwo, et al., Protein Science (1993), v 2, pp1715-1731; Liwo, et al., J. Comp. Chem. (1997), v 18, pp849-873; Liwo, et al., J. Comp. Chem. (1997), v 18, pp874-884; Liwo, et al., J. Comp. Chem. (1998), v 19, pp259-276; Forcefield for Protein Structure Prediction (Liwo, et al., Proc. Natl. Acad. Sci. USA (1999), v 96, pp5482-5485); ECEPP/3 (Liwo et al., J Protein Chem 1994 May;13(4):375-80); AMBER 1.1 force field (Weiner, et al., J. Am. Chem. Soc. v106, pp765-784); AMBER 3.0 force field (U.C. Singh et al., Proc. Natl. Acad. Sci. USA. 82:755-759); CHARMM and CHARMM22 (Brooks, et al., J. Comp. Chem. v4, pp 187-217); cvff3.0 (Dauber-Osguthorpe, et al., (1988) Proteins: Structure, Function and Genetics, v4, pp31-47); cff91 (Maple, et al., J. Comp. Chem. v15, 162-182); also, the DISCOVER (cvff and cff91) and AMBER forcefields are used in the INSIGHT molecular modeling package (Biosym/MSI, San Diego California) and HARMM is used in the QUANTA molecular modeling package (Biosym/MSI, San Diego California).

In addition, as outlined herein, a preferred method of generating a probability distribution table is through the use of sequence alignment programs. In addition, the probability table can be obtained by a combination of sequence alignments and computational approaches. For example, one can add amino acids found in the alignment of homologous sequences to the result of the computation. Preferable one can add the wild type amino acid identity to the probability table if it is not found in the computation.

As will be appreciated, a variant TR library created by recombining variable positions and/or residues at the variable position may not be in a rank-ordered list. In some embodiments, the entire list may just be made and tested. Alternatively, in a preferred embodiment, the variant TR library is also in the form of a rank ordered list. This may be done for several reasons, including the size of the library is still too big to generate experimentally, or for predictive purposes. This may be done in several ways.

WO 02/090300

PCT/US02/14358

In one embodiment, the library is ranked using the scoring functions of PDA to rank the library members. Alternatively, statistical methods could be used. For example, the library may be ranked by frequency score; that is, proteins containing the most of high frequency residues could be ranked higher, etc. This may be done by adding or multiplying the frequency at each variable position to generate a numerical score. Similarly, the library different positions could be weighted and then the proteins scored; for example, those containing certain residues could be arbitrarily ranked.

In a preferred embodiment, the different protein members of the variant TR library may be chemically synthesized. This is particularly useful when the designed proteins are short, preferably less than 150 amino acids in length, with less than 100 amino acids being preferred, and less than 50 amino acids being particularly preferred, although as is known in the art, longer proteins can be made chemically or enzymatically. See for example Wilken et al, Curr. Opin. Biotechnol. 9:412-26 (1998), hereby expressly incorporated by reference.

In a preferred embodiment, particularly for longer proteins or proteins for which large samples are desired, the library sequences are used to create nucleic acids such as DNA which encode the member sequences and which can then be cloned into host cells, expressed and assayed, if desired. Thus, nucleic acids, and particularly DNA, can be made which encodes each member protein sequence. This is done using well known procedures. The choice of codons, suitable expression vectors and suitable host cells will vary depending on a number of factors, and can be easily optimized as needed.

In a preferred embodiment, multiple PCR reactions with pooled oligonucleotides is done, as is generally described in U.S.S.N.09/927,790; incorporated herein by reference. In this embodiment, overlapping oligonucleotides are synthesized which correspond to the full length gene. Again, these oligonucleotides may represent all of the different amino acids at each variant position or subsets.

In a preferred embodiment, these oligonucleotides are pooled in equal proportions and multiple PCR reactions are performed to create full length sequences containing the combinations of mutations defined by the library. In addition, this may be done using error-prone PCR methods.

In a preferred embodiment, the different oligonucleotides are added in relative amounts corresponding to the probability distribution table. The multiple PCR reactions thus result in full length sequences with the desired combinations of mutations in the desired proportions.

The total number of oligonucleotides needed is a function of the number of positions being mutated and the number of mutations being considered at these positions:
 (number of oligos for constant positions) + M1 + M2 + M3 + ... Mn = (total number of oligos required),
 where Mn is the number of mutations considered at position n in the sequence.

WO 02/090300

PCT/US02/14358

In a preferred embodiment, each overlapping oligonucleotide comprises only one position to be varied; in alternate embodiments, the variant positions are too close together to allow this and multiple variants per oligonucleotide are used to allow complete recombination of all the possibilities. That is, each oligo can contain the codon for a single position being mutated, or for more than one position being mutated. The multiple positions being mutated must be close in sequence to prevent the oligo length from being impractical. For multiple mutating positions on an oligonucleotide, particular combinations of mutations can be included or excluded in the library by including or excluding the oligonucleotide encoding that combination. For example, as discussed herein, there may be correlations between variable regions; that is, when position X is a certain residue, position Y must (or must not) be a particular residue. These sets of variable positions are sometimes referred to herein as a "cluster". When the clusters are comprised of residues close together, and thus can reside on one oligonucleotide primer, the clusters can be set to the "good" correlations, and eliminate the bad combinations that may decrease the effectiveness of the library. However, if the residues of the cluster are far apart in sequence, and thus will reside on different oligonucleotides for synthesis, it may be desirable to either set the residues to the "good" correlation, or eliminate them as variable residues entirely. In an alternative embodiment, the library may be generated in several steps, so that the cluster mutations only appear together. This procedure, i.e. the procedure of identifying mutation clusters and either placing them on the same oligonucleotides or eliminating them from the library or library generation in several steps preserving clusters, can considerably enrich the experimental library with properly folded protein. Identification of clusters can be carried out by a number of ways, e.g. by using known pattern recognition methods, comparisons of frequencies of occurrence of mutations or by using energy analysis of the sequences to be experimentally generated (for example, if the energy of interaction is high, the positions are correlated). These correlations may be positional correlations (e.g. variable positions 1 and 2 always change together or never change together) or sequence correlations (e.g. if there is residue A at position 1, there is always residue B at position 2). See: Pattern discovery in Biomolecular Data: Tools, Techniques, and Applications; edited by Jason T.L. Wang, Bruce A. Shapiro, Dennis Shasha. New York: Oxford University, 1999; Andrews, Harry C. Introduction to mathematical techniques in pattern recognition; New York, Wiley-Interscience [1972]; Applications of Pattern Recognition; Editor, K.S. Fu. Boca Raton, Fla. CRC Press, 1982; Genetic Algorithms for Pattern Recognition; edited by Sankar K. Pal, Paul P. Wang. Boca Raton: CRC Press, c1996; Pandya, Abhijit S., Pattern recognition with neural networks in C++ / Abhijit S. Pandya, Robert B. Macy. Boca Raton, Fla.: CRC Press, 1996; Handbook of pattern recognition & computer vision / edited by C.H. Chen, L.F. Pau, P.S.P. Wang. 2nd ed. Singapore; River Edge, N.J.: World Scientific, c1999; Friedman, Introduction to Pattern Recognition: Statistical, Structural, Neural, and Fuzzy Logic Approaches; River Edge, N.J.: World Scientific, c1999, Series title: Series in machine perception and artificial intelligence; vol. 32; all of which are expressly incorporated by reference. In addition, programs used to search for consensus motifs can be used as well.

WO 02/090300

PCT/US02/14358

In addition, correlations and shuffling can be fixed or optimized by altering the design of the oligonucleotides; that is, by deciding where the oligonucleotides (primers) start and stop (e.g. where the sequences are "cut"). The start and stop sites of oligos can be set to maximize the number of clusters that appear in single oligonucleotides, thereby enriching the library with higher scoring sequences. Different oligonucleotide start and stop site options can be computationally modeled and ranked according to number of clusters that are represented on single oligos, or the percentage of the resulting sequences consistent with the predicted library of sequences.

The total number of oligonucleotides required increases when multiple mutable positions are encoded by a single oligonucleotide. The annealed regions are the ones that remain constant, i.e. have the sequence of the reference sequence.

Oligonucleotides with insertions or deletions of codons can be used to create a library expressing different length proteins. In particular computational sequence screening for insertions or deletions can result in secondary libraries defining different length proteins, which can be expressed by a library of pooled oligonucleotide of different lengths.

In a preferred embodiment, the variant TR library is done by shuffling the family (e.g. a set of variants); that is, some set of the top sequences (if a rank-ordered list is used) can be shuffled, either with or without error-prone PCR. "Shuffling" in this context means a recombination of related sequences, generally in a random way. It can include "shuffling" as defined and exemplified in U.S. Patent Nos. 5,830,721; 5,811,238; 5,605,793; 5,837,458 and PCT US/19256, all of which are expressly incorporated by reference in their entirety. This set of sequences can also be an artificial set; for example, from a probability table (for example generated using SCMF) or a Monte Carlo set. Similarly, the "family" can be the top 10 and the bottom 10 sequences, the top 100 sequence, etc. This may also be done using error-prone PCR.

Thus, in a preferred embodiment, in silico shuffling is done using the computational methods described herein. That is, starting with either two libraries or two sequences, random recombinations of the sequences can be generated and evaluated.

In a preferred embodiment, error-prone PCR is done to generate the variant TR library. See U.S. Patent Nos. 5,605,793, 5,811,238, and 5,830,721, all of which are hereby incorporated by reference. This can be done on the optimal sequence or on top members of the library, or some other artificial set or family. In this embodiment, the gene for the optimal sequence found in the computational screen of the primary library can be synthesized. Error prone PCR is then performed on the optimal sequence gene in the presence of oligonucleotides that code for the mutations at the variant positions of the library (bias oligonucleotides). The addition of the oligonucleotides will create a bias favoring

WO 02/090300

PCT/US02/14358

the incorporation of the mutations in the library. Alternatively, only oligonucleotides for certain mutations may be used to bias the library.

- In a preferred embodiment, gene shuffling with error prone PCR can be performed on the gene for the optimal sequence, in the presence of bias oligonucleotides, to create a DNA sequence library that reflects the proportion of the mutations found in the variant TR library. The choice of the bias oligonucleotides can be done in a variety of ways; they can be chosen on the basis of their frequency, i.e. oligonucleotides encoding high mutational frequency positions can be used; alternatively, oligonucleotides containing the most variable positions can be used, such that the diversity is increased; if the secondary library is ranked, some number of top scoring positions can be used to generate bias oligonucleotides; random positions may be chosen; a few top scoring and a few low scoring ones may be chosen; etc. What is important is to generate new sequences based on preferred variable positions and sequences.
- In a preferred embodiment, PCR using a wild type gene or other gene can be used, as is generally described in U.S.S.N. 09/927,790; incorporated herein by reference. In this embodiment, a starting gene is used; generally, although this is not required, the gene is usually the wild type gene. In some cases it may be the gene encoding the global optimized sequence, or any other sequence of the list, or a consensus sequence obtained e.g. from aligning homologous sequences from different organisms. In this embodiment, oligonucleotides are used that correspond to the variant positions and contain the different amino acids of the library. PCR is done using PCR primers at the termini, as is known in the art. This provides two benefits; the first is that this generally requires fewer oligonucleotides and can result in fewer errors. In addition, it has experimental advantages in that if the wild type gene is used, it need not be synthesized.
- In addition, there are several other techniques that can be used, as exemplified in the figures. In a preferred embodiment, ligation of PCR products is done.
- In a preferred embodiment, a variety of additional steps may be done to the variant TR library; for example, further computational processing can occur, different variant TR libraries can be recombined, or cutoffs from different libraries can be combined. In a preferred embodiment, a variant TR library may be computationally remanipulated to form an additional variant TR library (sometimes referred to herein as "tertiary libraries"). For example, any of the variant TR library sequences may be chosen for a second round of PDA, by freezing or fixing some or all of the changed positions in the first library. Alternatively, only changes seen in the last probability distribution table are allowed. Alternatively, the stringency of the probability table may be altered, either by increasing or decreasing the cutoff for inclusion. Similarly, the variant TR library may be recombined experimentally after the first round; for example, the best gene/genes from the first screen may be taken and gene assembly redone (using techniques outlined below, multiple PCR, error prone PCR, shuffling, etc.).

WO 02/090300

PCT/US02/14358

Alternatively, the fragments from one or more good gene(s) to change probabilities at some positions. This biases the search to an area of sequence space found in the first round of computational and experimental screening.

- 5 In a preferred embodiment, a tertiary library can be generated from combining different variant TR-libraries. For example, a probability distribution table from a first variant TR library can be generated and recombined, either computationally or experimentally, as outlined herein. A PDA variant TR library may be combined with a sequence alignment variant TR library, and either recombined (again, computationally or experimentally) or just the cutoffs from each joined to make a new tertiary library.
- 10 The top sequences from several libraries can be recombined. Sequences from the top of a library can be combined with sequences from the bottom of the library to more broadly sample sequence space, or only sequences distant from the top of the library can be combined. Variant TR libraries that analyzed different parts of a protein can be combined to a tertiary library that treats the combined parts of the protein.
- 15 In a preferred embodiment, a tertiary library can be generated using correlations in a variant TR library. That is, a residue at a first variable position may be correlated to a residue at second variable position (or correlated to residues at additional positions as well). For example, two variable positions may sterically or electrostatically interact, such that if the first residue is X, the second residue must be Y. This may be either a positive or negative correlation.
- 20

Using the nucleic acids of the present invention that encode candidate variant proteins or candidate variant library members, a variety of expression vectors are made. The expression vectors may be either self-replicating extrachromosomal vectors or vectors which integrate into a host genome.

- 25 Generally, these expression vectors include transcriptional and translational regulatory nucleic acid operably linked to the nucleic acid encoding the library protein. The term "control sequences" refers to DNA sequences necessary for the expression of an operably linked coding sequence in a particular host organism. The control sequences that are suitable for prokaryotes, for example, include a promoter, optionally an operator sequence, and a ribosome binding site. Eukaryotic cells are known to utilize promoters, polyadenylation signals, and enhancers.
- 30

- Nucleic acid is "operably linked" when it is placed into a functional relationship with another nucleic acid sequence. For example, DNA for a presequence or secretory leader is operably linked to DNA for a polypeptide if it is expressed as a preprotein that participates in the secretion of the polypeptide; a promoter or enhancer is operably linked to a coding sequence if it affects the transcription of the sequence; or a ribosome binding site is operably linked to a coding sequence if it is positioned so as to facilitate translation. Generally, "operably linked" means that the DNA sequences being linked are contiguous, and, in the case of a secretory leader, contiguous and in reading phase. However, enhancers do not have to be contiguous. Linking is accomplished by ligation at convenient restriction
- 35

WO 02/090300

PCT/US02/14358

sites. If such sites do not exist, the synthetic oligonucleotide adaptors or linkers are used in accordance with conventional practice. The transcriptional and translational regulatory nucleic acid will generally be appropriate to the host cell used to express the library protein, as will be appreciated by those in the art; for example, transcriptional and translational regulatory nucleic acid sequences from *Bacillus* are preferably used to express the library protein in *Bacillus*. Numerous types of appropriate expression vectors, and suitable regulatory sequences are known in the art for a variety of host cells.

In general, the transcriptional and translational regulatory sequences may include, but are not limited to, promoter sequences, ribosomal binding sites, transcriptional start and stop sequences, translational start and stop sequences, and enhancer or activator sequences. In a preferred embodiment, the regulatory sequences include a promoter and transcriptional start and stop sequences.

Promoter sequences include constitutive and inducible promoter sequences. The promoters may be either naturally occurring promoters, hybrid or synthetic promoters. Hybrid promoters, which combine elements of more than one promoter, are also known in the art, and are useful in the present invention.

In addition, the expression vector may comprise additional elements. For example, the expression vector may have two replication systems, thus allowing it to be maintained in two organisms, for example in mammalian or insect cells for expression and in a prokaryotic host for cloning and amplification. Furthermore, for integrating expression vectors, the expression vector contains at least one sequence homologous to the host cell genome, and preferably two homologous sequences which flank the expression construct. The integrating vector may be directed to a specific locus in the host cell by selecting the appropriate homologous sequence for inclusion in the vector. Constructs for integrating vectors and appropriate selection and screening protocols are well known in the art and are described in e.g., Mansour et al., *Cell*, 51:503 (1988) and Murray, *Gene Transfer and Expression Protocols, Methods in Molecular Biology, Vol. 7* (Clifton: Humana Press, 1991).

In addition, in a preferred embodiment, the expression vector contains a selection gene to allow the selection of transformed host cells containing the expression vector, and particularly in the case of mammalian cells, ensures the stability of the vector, since cells which do not contain the vector will generally die. Selection genes are well known in the art and will vary with the host cell used. By "selection gene" herein is meant any gene which encodes a gene product that confers resistance to a selection agent. Suitable selection agents include, but are not limited to, neomycin (or its analog G418), blasticidin S, histidinol D, bleomycin, puromycin, hygromycin B, and other drugs.

In a preferred embodiment, the expression vector contains a RNA splicing sequence upstream or downstream of the gene to be expressed in order to increase the level of gene expression. See

WO 02/090300

PCT/US02/14358

Barret et al., Nucleic Acids Res. 1991; Groos et al., Mol. Cell. Biol. 1987; and Budiman et al., Mol. Cell. Biol. 1988.

A preferred expression vector system is a retroviral vector system such as is generally described in
 5 Mann et al., Cell, 33:153-9 (1993); Pear et al., Proc. Natl. Acad. Sci. U.S.A., 90(16):8392-6 (1993);
 Kitamura et al., Proc. Natl. Acad. Sci. U.S.A., 92:9146-50 (1995); Kinsella et al., Human Gene
 Therapy, 7:1405-13; Hofmann et al., Proc. Natl. Acad. Sci. U.S.A., 93:5185-90; Choate et al., Human
 Gene Therapy, 7:2247 (1996); PCT/US97/01019 and PCT/US97/01048, and references cited therein,
 all of which are hereby expressly incorporated by reference.

10 The candidate variant library proteins of the present invention are produced by culturing a host cell
 transformed with nucleic acid, preferably an expression vector, containing nucleic acid encoding an
 library protein, under the appropriate conditions to induce or cause expression of the library protein.
 The conditions appropriate for candidate variant library protein expression will vary with the choice of
 15 the expression vector and the host cell, and will be easily ascertained by one skilled in the art through
 routine experimentation. For example, the use of constitutive promoters in the expression vector will
 require optimizing the growth and proliferation of the host cell, while the use of an inducible promoter
 requires the appropriate growth conditions for induction. In addition, in some embodiments, the timing
 of the harvest is important. For example, the baculoviral systems used in insect cell expression are
 20 lytic viruses, and thus harvest time selection can be crucial for product yield.

As will be appreciated by those in the art, the type of cells used in the present invention can vary
 widely. Basically, a wide variety of appropriate host cells can be used, including yeast, bacteria,
 archaeobacteria, fungi, and insect, plant, and animal cells, including mammalian cells. Of particular
 25 interest are *Drosophila melanogaster* cells, *Saccharomyces cerevisiae* and other yeasts, *E. coli*,
Bacillus subtilis, SF9 cells, C129 cells, 293 cells, Neurospora, BHK, CHO, COS, and HeLa cells,
 fibroblasts, Schwannoma cell lines, immortalized mammalian myeloid and lymphoid cell lines, Jurkat
 cells, mast cells and other endocrine and exocrine cells, and neuronal cells. See the ATCC cell line
 catalog, hereby expressly incorporated by reference. In addition, the expression of the secondary
 30 libraries in phage display systems, such as are well known in the art, are particularly preferred,
 especially when the secondary library comprises random peptides. In one embodiment, the cells may
 be genetically engineered, that is, contain exogenous nucleic acid, for example, to contain target
 molecules.

35 In a preferred embodiment, the candidate variant protein or candidate variant library proteins are
 expressed in mammalian cells. Any mammalian cells may be used, with mouse, rat, primate and
 human cells being particularly preferred, although as will be appreciated by those in the art,
 modifications of the system by pseudotyping allows all eukaryotic cells to be used, preferably higher
 eukaryotes. As is more fully described below, a screen will be set up such that the cells exhibit a

WO 02/090300

PCT/US02/14358

selectable phenotype in the presence of a random library member. As is more fully described below, cell types implicated in a wide variety of disease conditions are particularly useful, so long as a suitable screen may be designed to allow the selection of cells that exhibit an altered phenotype as a consequence of the presence of a library member within the cell.

5

Accordingly, suitable mammalian cell types include, but are not limited to, tumor cells of all types (particularly melanoma, myeloid leukemia, carcinomas of the lung, breast, ovaries, colon, kidney, prostate, pancreas and testes), cardiomyocytes, endothelial cells, epithelial cells, lymphocytes (T-cell and B cell), mast cells, eosinophils, vascular intimal cells, hepatocytes, leukocytes including mononuclear leukocytes, stem cells such as haemopoietic, neural, skin, lung, kidney, liver and myocyte stem cells (for use in screening for differentiation and de-differentiation factors), osteoclasts, chondrocytes and other connective tissue cells, keratinocytes, melanocytes, liver cells, kidney cells, and adipocytes. Suitable cells also include known research cells, including, but not limited to, Jurkat T cells, NIH3T3 cells, CHO, Cos, etc. See the ATCC cell line catalog, hereby expressly incorporated by reference.

15

Mammalian expression systems are also known in the art, and include retroviral systems. A mammalian promoter is any DNA sequence capable of binding mammalian RNA polymerase and initiating the downstream (3') transcription of a coding sequence for library protein into mRNA. A promoter will have a transcription initiating region, which is usually placed proximal to the 5' end of the coding sequence, and a TATA box, usually located 25-30 base pairs upstream of the transcription initiation site. The TATA box is thought to direct RNA polymerase II to begin RNA synthesis at the correct site. A mammalian promoter will also contain an upstream promoter element (enhancer element), typically located within 100 to 200 base pairs upstream of the TATA box. An upstream promoter element determines the rate at which transcription is initiated and can act in either orientation. Of particular use as mammalian promoters are the promoters from mammalian viral genes, since the viral genes are often highly expressed and have a broad host range. Examples include the SV40 early promoter, mouse mammary tumor virus LTR promoter, adenovirus major late promoter, herpes simplex virus promoter, and the CMV promoter.

20

25

30

Typically, transcription termination and polyadenylation sequences recognized by mammalian cells are regulatory regions located 3' to the translation stop codon and thus, together with the promoter elements, flank the coding sequence. The 3' terminus of the mature mRNA is formed by site-specific post-translational cleavage and polyadenylation. Examples of transcription terminator and polyadenylation signals include those derived from SV40.

35

The methods of introducing exogenous nucleic acid into mammalian hosts, as well as other hosts, is well known in the art, and will vary with the host cell used. Techniques include dextran-mediated transfection, calcium phosphate precipitation, polybrene mediated transfection, protoplast fusion,

WO 02/090300

PCT/US02/14358

electroporation, viral infection, encapsulation of the polynucleotide(s) in liposomes, and direct microinjection of the DNA into nuclei.

5 In a preferred embodiment, candidate variant proteins or candidate variant library proteins are expressed in bacterial systems. Bacterial expression systems are well known in the art.

10 A suitable bacterial promoter is any nucleic acid sequence capable of binding bacterial RNA polymerase and initiating the downstream (3') transcription of the coding sequence of library protein into mRNA. A bacterial promoter has a transcription initiation region which is usually placed proximal to the 5' end of the coding sequence. This transcription initiation region typically includes an RNA polymerase binding site and a transcription initiation site. Sequences encoding metabolic pathway enzymes provide particularly useful promoter sequences. Examples include promoter sequences derived from sugar metabolizing enzymes, such as galactose, lactose and maltose, and sequences derived from biosynthetic enzymes such as tryptophan. Promoters from bacteriophage may also be used and are known in the art. In addition, synthetic promoters and hybrid promoters are also useful; for example, the *tac* promoter is a hybrid of the *trp* and *lac* promoter sequences. Furthermore, a bacterial promoter can include naturally occurring promoters of non-bacterial origin that have the ability to bind bacterial RNA polymerase and initiate transcription.

15 In addition to a functioning promoter sequence, an efficient ribosome binding site is desirable. In *E. coli*, the ribosome binding site is called the Shine-Delgarno (SD) sequence and includes an initiation codon and a sequence 3-9 nucleotides in length located 3 - 11 nucleotides upstream of the initiation codon.

20

25 The expression vector may also include a signal peptide sequence that provides for secretion of the library protein in bacteria. The signal sequence typically encodes a signal peptide comprised of hydrophobic amino acids which direct the secretion of the protein from the cell, as is well known in the art. The protein is either secreted into the growth media (gram-positive bacteria) or into the periplasmic space, located between the inner and outer membrane of the cell (gram-negative bacteria).

30 The bacterial expression vector may also include a selectable marker gene to allow for the selection of bacterial strains that have been transformed. Suitable selection genes include genes which render the bacteria resistant to drugs such as ampicillin, chloramphenicol, erythromycin, kanamycin, neomycin and tetracycline. Selectable markers also include biosynthetic genes, such as those in the histidine, tryptophan and leucine biosynthetic pathways.

35

These components are assembled into expression vectors. Expression vectors for bacteria are well known in the art, and include vectors for *Bacillus subtilis*, *E. coli*, *Streptococcus cremoris*, and *Streptococcus lividans*, among others.

WO 02/090300

PCT/US02/14358

The bacterial expression vectors are transformed into bacterial host cells using techniques well known in the art, such as calcium chloride treatment, electroporation, and others.

- 5 In one embodiment, candidate variant protein are produced in insect cells. Expression vectors for the transformation of insect cells, and in particular, baculovirus-based expression vectors, are well known in the art and are described e.g., in O'Reilly et al., *Baculovirus Expression Vectors: A Laboratory Manual* (New York: Oxford University Press, 1994).
- 10 In a preferred embodiment, candidate variant protein is produced in yeast cells. Yeast expression systems are well known in the art, and include expression vectors for *Saccharomyces cerevisiae*, *Candida albicans* and *C. maltosa*, *Hansenula polymorpha*, *Kluyveromyces fragilis* and *K. lactis*, *Pichia guilliermondii* and *P. pastoris*, *Schizosaccharomyces pombe*, and *Yarrowia lipolytica*. Preferred promoter sequences for expression in yeast include the inducible GAL1,10 promoter, the promoters from alcohol dehydrogenase, enolase, glucokinase, glucose-6-phosphate isomerase, glyceraldehyde-3-phosphate-dehydrogenase, hexokinase, phosphofructokinase, 3-phosphoglycerate mutase, pyruvate kinase, and the acid phosphatase gene. Yeast selectable markers include ADE2, HIS4, LEU2, TRP1, and ALG7, which confers resistance to tunicamycin; the neomycin phosphotransferase gene, which confers resistance to G418; and the CUP1 gene, which allows yeast to grow in the
- 20 presence of copper ions.

- In a preferred embodiment, the candidate variant protein or candidate variant library proteins are expressed in plant cells. Gene sequences intended for expression in transgenic plants are first assembled in expression cassettes adjacent to a suitable promoter expressible in plants. The
- 25 expression cassettes may also include any further sequences required or selected for the expression of the transgene. Such sequences include, but are not restricted to, transcription terminators, extraneous sequences to enhance expression such as introns, enhancer sequences, and sequences intended for the targeting of the gene product to specific organelles and cell compartments. These expression cassettes can then be easily transferred to the plant transformation vectors described
- 30 below. The following is a description of various components of typical expression cassettes.

- The selection of the promoter used in expression cassettes determines the spatial and temporal expression pattern of the transgene in the transgenic plant. Selected promoters express transgenes in specific cell types (such as leaf epidermal cells, mesophyll cells, root cortex cells) or in specific
- 35 tissues or organs (roots, leaves or flowers, for example) and the selection of a promoter is therefore based on the desired location of accumulation of the gene product. In a preferred embodiment of the invention, a seed-specific promoter is used for expression of an oleosin-TR fusion protein, an oleosin-TR fusion protein or an oleosin-hybrid TR/TR-reductase fusion protein. In a most preferred embodiment, the seed specific promoter is a phaseolin promoter.

WO 02/090300

PCT/US02/14358

Promoters vary in their ability to promote transcription. Depending upon the host cell system utilized, any one of a number of suitable promoters known in the art can be used. For constitutive expression, the CaMV 35S promoter, the rice actin promoter, or the ubiquitin promoter may be used.

- 5 Alternatively, an inducible promoter may be selected to drive expression of the gene under various inducing conditions. For chemically inducible expression, the inducible PR-1 promoter from tobacco or *Arabidopsis* may be used (see, e.g., U.S. Patent No. 5,689,044).

- 10 A variety of transcriptional terminators are available for use in nuclear gene expression cassettes, and are responsible for the termination of transcription beyond the transgene and its correct polyadenylation. Appropriate transcriptional terminators are those that are known to function in plants and include the CaMV 35S terminator, the tm¹ terminator, the nopaline synthase (nos) terminator and the pea rbcS E9 terminator. These can be used in both monocotyledonous and dicotyledonous plants. In a preferred embodiment, a phasedolin transcriptional terminator is used. Expression in
15 plastids may not require termination, but may require correct 5' and 3' signals for translational initiation, elongation and RNA stability.

- 20 Numerous sequences have been found to enhance gene expression from within the transcriptional unit and these sequences can be used in conjunction with the genes of this invention to increase their expression in transgenic plants. For example, various intron sequences such as introns of the maize Adh1 gene have been shown to enhance expression, particularly in monocotyledonous cells. In addition, a number of non-translated leader sequences derived from viruses are also known to enhance expression, and these are particularly effective in dicotyledonous cells.

- 25 For their expression in transgenic plants, the coding sequence of DNA molecules used may require modification and optimization, particularly when the DNA molecules are of prokaryotic origin. It is known in the art that all organisms have specific preferences for codon usage, and the codons in the nucleotide sequence of the DNA molecules of the present invention can be changed to conform with specific plant preferences, while maintaining the amino acids encoded thereby. High expression in plants is best achieved from coding sequences which have at least 35% GC content, and preferably
30 more than 45%. Nucleotide sequences which have low GC contents may express poorly due to the existence of ATTTA motifs which may destabilize messages, and AATAAA motifs which may cause inappropriate polyadenylation. Although preferred gene sequences may be adequately expressed in both monocotyledonous and dicotyledonous plant species, sequences can be modified to account for the specific codon preferences and GC content preferences of monocotyledons or dicotyledons as
35 these preferences have been shown to differ (Murray *et al.* (1989) *Nucl Acids Res* 17: 477-498). In addition, the nucleotide sequences are screened for the existence of illegitimate splice sites which cause message truncation. All changes required to be made within the nucleotide sequences such as those described above are made using well known techniques of site directed mutagenesis, PCR, and synthetic gene construction using, for example, the methods described in the published patent

WO 02/090300

PCT/US02/14358

applications EP 0 385 962, EP 0 359 472, and WO 93/07278, the entire disclosures of which are hereby incorporated in their entireties.

For efficient initiation of translation, sequences adjacent to the initiating methionine may require modification. For example, they can be modified by the inclusion of sequences known to be effective in plants. Joshi has suggested an appropriate consensus for plants (*Nuc Acids Res* (1987) 15:6643-6653) and a further consensus translation initiator (*Clontech* 1993/1994 catalog, page 210) may be included. These consensus sequences are suitable for use with the nucleotide sequences of this invention. The sequences are incorporated into constructions including the nucleotide sequence, up to and including the ATG (whilst leaving the second amino acid unmodified), or alternatively up to and including the GTC subsequent to the ATG (with the possibility of modifying the second amino acid of the transgene).

Various mechanisms for targeting gene products are known to exist in plants, and the sequences controlling the functioning of these mechanisms have been characterized in some detail. For example, the targeting of gene products to the chloroplast is controlled by a transit sequence found at the amino terminal end of various proteins which is cleaved during chloroplast import to yield the mature protein (Comai *et al.* (1988) *J Biol Chem* 263: 15104-15109). Other gene products are localized to other organelles such as the mitochondrion and the peroxisome (Unger *et al.* (1989) *Plant Mol Biol* 13:411-418). The cDNAs encoding these products can be manipulated to target heterologous gene products to these organelles. In addition, sequences have been characterized which cause the targeting of gene products to other cell compartments.

Amino terminal sequences are responsible for targeting to the ER, the apoplast, and extracellular secretion from aleurone cells (Koehler & Ho (1990) *Plant Cell* 2:769-783). Additionally, amino terminal sequences in conjunction with carboxy terminal sequences are responsible for vacuolar targeting of gene products (Shinshi *et al.*, (1990) *Plant Mol Biol* 14:357-368). By the fusion of the appropriate targeting sequences described above to transgene sequences of interest it is possible to direct the transgene product to the desired organelle or cell compartment.

In another preferred embodiment, the DNA molecules of this invention are directly transformed into the plastid genome. Plastid transformation technology is described extensively in U.S. Patent Nos. 5,451,513, 5,545,817, 5,545,818 and 5,576,198; in PCT application nos. WO 95/16783 and WO 97/32977; and in McBride *et al.*, *Proc Natl Acad Sci USA* 91: 7301-7305 (1994), the entire disclosures of all of which are hereby incorporated by reference. In one embodiment, plastid transformation is achieved via biolistics, first carried out in the unicellular green alga *Chlamydomonas reinhardtii* (Boynton *et al.* (1988) *Science* 240:1534-1537)) and then extended to *Nicotiana tabacum* (Svab *et al.* (1980) *Proc Natl Acad Sci USA* 87:8526-8530), combined with selection for cis-acting

WO 02/090300

PCT/US02/14358

antibiotic resistance loci (spectinomycin or streptomycin resistance) or complementation of non-photosynthetic mutant phenotypes.

In other embodiment, tobacco plastid transformation is carried out by particle bombardment of leaf or callus tissue, or polyethylene glycol (PEG)-mediated uptake of plasmid DNA by protoplasts, using cloned plastid DNA flanking a selectable antibiotic resistance marker. The 1 to 1.5 kb flanking regions, termed targeting sequences, facilitate homologous recombination with the plastid genome and allow the replacement or modification of specific regions of the 156 kb tobacco plastid genome. Initially, point mutations in the plastid 16S rDNA and rps12 genes conferring resistance to spectinomycin and/or streptomycin were utilized as selectable markers for transformation (Svab *et al.* (1990) *Proc Natl Acad Sci USA* 87:8526-8530; Staub *et al.* (1992) *Plant Cell* 4:39-45, the entire disclosures of which are hereby incorporated by reference), resulting in stable homoplasmic transformants at a frequency of approximately one per 100 bombardments of target leaves. The presence of cloning sites between these markers allows creation of a plastid targeting vector for introduction of foreign genes (Staub *et al.* (1993) *EMBO J* 12:601-606, the entire disclosure of which is hereby incorporated by reference). Substantial increases in transformation frequency were obtained by replacement of the recessive rRNA or r-protein antibiotic resistance genes with a dominant selectable marker, the bacterial *aadA* gene encoding the spectinomycin-detoxifying enzyme aminoglycoside-3'-adenyltransferase (Svab *et al.* (1993) *Proc Natl Acad Sci USA* 90: 913-917, the entire disclosure of which is hereby incorporated by reference). Previously, this marker had been used successfully for high-frequency transformation of the plastid genome of the green alga *Chlamydomonas reinhardtii* (Goldschmidt-Clermont, M. (1991) *Nucl Acids Res* 19, 4083-4089, the entire disclosure of which is hereby incorporated by reference). Recently, plastid transformation of protoplasts from tobacco and the moss *Physcomitrella* has been attained using PEG-mediated DNA uptake (O'Neill *et al.* (1993) *Plant J* 3:729-738; Koop *et al.* (1996) *Planta* 199:193-201, the entire disclosures of which are hereby incorporated by reference).

Both particle bombardment and protoplast transformation are appropriate in the context of the present invention. Plastid transformation of oilseed plants has been successfully carried out in the genera *Arabidopsis* and *Brassica* (Sikdar *et al.* (1998) *Plant Cell Rep* 18:20-24; PCT Application WO 00/39313, the entire disclosures of which are hereby incorporated by reference).

A DNA molecule of the present invention is inserted into a plastid expression cassette including a promoter capable of expressing the DNA molecule in plant plastids. A preferred promoter capable of expression in a plant plastid is, for example, a promoter isolated from the 5' flanking region upstream of the coding region of a plastid gene, which may come from the same or a different species, and the native product of which is typically found in a majority of plastid types including those present in non-green tissues. Gene expression in plastids differs from nuclear gene expression and is related to

WO 02/090300

PCT/US02/14358

gene expression in prokaryotes (Stern *et al.* (1997) *Trends in Plant Sci* 2:308-315, the entire disclosure of which is hereby incorporated by reference).

- Plastid promoters generally contain the -35 and -10 elements typical of prokaryotic promoters, and some plastid promoters called PEP (plastid-encoded RNA polymerase) promoters are recognized by an *E. coli*-like RNA polymerase mostly encoded in the plastid genome, while other plastid promoters called NEP promoters are recognized by a nuclear-encoded RNA polymerase. Both types of plastid promoters are suitable for the present invention. Examples of plastid promoters include promoters of *clpP* genes such as the tobacco *clpP* gene promoter (WO 97/06250, the entire disclosure of which is hereby incorporated by reference) and the *Arabidopsis clpP* gene promoter (U.S. Application No. 09/038,878, the entire disclosure of which is hereby incorporated by reference). Another promoter capable of driving expression of a DNA molecule in plant plastids comes from the regulatory region of the plastid 16S ribosomal RNA operon (Harris *et al.*, (1994) *Microbiol Rev* 58:700-754; Shinzaki *et al.* (1986) *EMBO J* 5:2043-2049, the entire disclosures of both of which are hereby incorporated by reference). Other examples of promoters capable of driving expression of a DNA molecule in plant plastids include a *psbA* promoter or an *rbcl* promoter. A plastid expression cassette preferably further includes a plastid gene 3' untranslated sequence (3' UTR) operatively linked to a DNA molecule of the present invention. The role of untranslated sequences is preferably to direct the 3' processing of the transcribed RNA rather than termination of transcription. Preferably, the 3' UTR is a plastid *rps16* gene 3' untranslated sequence, or the *Arabidopsis* plastid *psbA* gene 3' untranslated sequence. In a further preferred embodiment, a plastid expression cassette includes a poly-G tract instead of a 3' untranslated sequence. A plastid expression cassette also preferably further includes a 5' untranslated sequence (5' UTR) functional in plant plastids, operatively linked to a DNA molecule of the present invention.
- A plastid expression cassette is included in a plastid transformation vector, which preferably further includes flanking regions for integration into the plastid genome by homologous recombination. The plastid transformation vector may optionally include at least one plastid origin of replication. The present invention also encompasses a plant plastid transformed with such a plastid transformation vector, wherein the DNA molecule is expressible in the plant plastid. The invention also encompasses a plant or plant cell, including the progeny thereof, including this plant plastid. In a preferred embodiment, the plant or plant cell, including the progeny thereof, is homoplasmic for transgenic plastids.
- Other promoters capable of driving expression of a DNA molecule in plant plastids include transactivator-regulated promoters, preferably heterologous with respect to the plant or to the subcellular organelle or component of the plant cell in which expression is effected. In these cases, the DNA molecule encoding the transactivator is inserted into an appropriate nuclear expression cassette which is transformed into the plant nuclear DNA. The transactivator is targeted to plastids

WO 02/090300

PCT/US02/14358

using a plastid transit peptide. The transactivator and the transactivator-driven DNA molecule are brought together either by crossing a selected plastid-transformed line with a transgenic line containing a DNA molecule encoding the transactivator supplemented with a plastid-targeting sequence and operably linked to a nuclear promoter, or by directly transforming a plastid transformation vector containing the desired DNA molecule into a transgenic line containing a DNA molecule encoding the transactivator supplemented with a plastid-targeting sequence operably linked to a nuclear promoter. If the nuclear promoter is an inducible promoter, in particular a chemically inducible promoter, expression of the DNA molecule in the plastids of plants is activated by foliar application of a chemical inducer. Such an inducible transactivator-mediated plastid expression system is preferably tightly regulatable, with no detectable expression prior to induction and exceptionally high expression and accumulation of protein following induction. A preferred transactivator is, for example, viral RNA polymerase. Preferred promoters of this type are promoters recognized by a single sub-unit RNA polymerase, such as the T7 gene 10 promoter, which is recognized by the bacteriophage T7 DNA-dependent RNA polymerase. The gene encoding the T7 polymerase is preferably transformed into the nuclear genome and the T7 polymerase is targeted to the plastids using a plastid transit peptide. Promoters suitable for nuclear expression of a gene, for example a gene encoding a viral RNA polymerase such as the T7 polymerase, are described above and elsewhere in this application. Expression of DNA molecules in plastids can be constitutive or can be inducible, and such plastid expression can be also organ- or tissue-specific. Examples of various expression systems are extensively described in WO 98/11235, the entire disclosure of which is hereby incorporated by reference. Thus, in one aspect, the present invention utilized coupled expression in the nuclear genome of a chloroplast-targeted phage T7 RNA polymerase under the control of the chemically inducible PR-1a promoter, for example of the PR-1 promoter of tobacco, operably linked with a chloroplast reporter transgene regulated by T7 gene 10 promoter/terminator sequences, for example as described in as in US Patent No. 5,614,395 the entire disclosure of which is hereby incorporated by reference. In another embodiment, when plastid transformants homoplasmic for the maternally inherited TR genes are pollinated by lines expressing the T7 polymerase in the nucleus, F1 plants are obtained that carry both transgene constructs but do not express them until synthesis of large amounts of enzymatically active protein in the plastids is triggered by foliar application of the PR-1a inducer compound benzo(1,2,3)thiadiazole-7-carbothioic acid S-methyl ester (BTH).

In a preferred embodiment, two or more genes, for example TR genes, are transcribed from the plastid genome from a single promoter in an operon-like polycistronic gene. In a preferred embodiment, the operon-like polycistronic gene includes an intervening DNA sequence between two genes in the operon-like polycistronic gene. In a preferred embodiment, the DNA sequence is not present in the plastid genome to avoid homologous recombination with plastid sequences. In another preferred embodiment, the DNA sequence is derived from the 5' untranslated (UTR) region of a non-eukaryotic gene, preferably from a viral 5'UTR, preferably from a 5'UTR derived from a bacterial

WO 02/090300

PCT/US02/14358

phage, such as a T7, T3 or SP6 phage. In a preferred embodiment, a portion of the DNA sequence may be modified to prevent the formation of RNA secondary structures in an RNA transcript of the operon-like polycistronic gene, for example between the DNA sequence and the RBS of the downstream gene. Such secondary structures may inhibit or repress the expression of the downstream gene, particularly the initiation of translation. Such RNA secondary structures are

5 predicted by determining their melting temperatures using computer models and programs such as the "mfold" program version 3 (available from Zuker and Turner, Washington University School of Medicine, St-Louis, MO) and other methods known to one skilled in the art.

10 The presence of the intervening DNA sequence in the operon-like polycistronic gene increases the accessibility of the RBS of the downstream gene, thus resulting in higher rates of expression. Such strategy is applicable to any two or more genes to be transcribed from the plastid genome from a single promoter in an operon-like chimeric gene.

15 Numerous transformation vectors available for plant transformation are known to those of ordinary skill in the art, and the genes pertinent to this invention can be used in conjunction with any such vectors. Vector selection will depend upon the preferred transformation technique and the target species being transformed. For certain target species, different antibiotic or herbicide selection markers may be preferred.

20 Selection markers used routinely in transformation include the *nptII* gene, which confers resistance to kanamycin and related antibiotics (Messing & Vieira. (1982) *Gene* 19:259-268; Bevan *et al.* (1983) *Nature* 304:184-187), the *bar* gene, which confers resistance to the herbicide phosphinothricin (White *et al.* (1990) *Nucl Acids Res* 18: 1062; Spencer *et al.* (1990) *Theor Appl Genet* 79:625-631), the *hph* gene, which confers resistance to the antibiotic hygromycin (Yanofsky, *et al.* (1992) *Gene* 117:161-167), the *dhfr* gene, which confers resistance to methotrexate (Bourouis *et al.*, *EMBO J.* 7:1099-1104 (1983)), the *EPSPS* gene, which confers resistance to glyphosate (U.S. Patent Nos. 4,940,935 and 5,188,642), and the mannose phosphate isomerase gene *pmi* which confers tolerance to normally phytotoxic sugar mannose (Negrotto, *et al.* (2000) *Plant Cell Rep* 19:798-803).

30 Many vectors are suitable for transformation using *Agrobacterium tumefaciens*. These typically carry at least one T-DNA border sequence and include vectors such as pBIN 19 (Bevan, (1984) *Nucl Acids Res*) and pXYZ. Typical vectors suitable for *Agrobacterium* transformation include the binary vectors pCIB200 and pCIB2001, as well as the binary vector pCIB1 0 and hygromycin selection derivatives thereof. (U.S. Patent No. 5,639,949).

35 Transformation without the use of *Agrobacterium tumefaciens* circumvents the requirement for T-DNA sequences in the chosen transformation vector. Consequently, vectors lacking these sequences can be used as an alternative to vectors such as the T-DNA-containing vectors described above.

WO 02/090300

PCT/US02/14358

Transformation techniques that do not rely on *Agrobacterium* include transformation via particle bombardment, protoplast uptake for example PEG and/or electroporation, and microinjection. The choice of vector depends largely on the preferred selection for the species being transformed. Typical vectors suitable for non-*Agrobacterium* transformation include pCIB3064, pSOG1 9, and pSOG35.

5 (U.S. Patent No. 5,639,949).

Once the coding sequence of interest has been cloned into an expression system, it is transformed into a plant cell. Methods for transformation and regeneration of plants are well known in the art. For example, Ti plasmid vectors have been utilized for the delivery of foreign DNA, as well as direct uptake of DNA, liposomes, electroporation, microinjection, and microprojectiles. In addition, bacteria from the genus *Agrobacterium* can be utilized to transform plant cells.

Transformation techniques for dicotyledons are well known in the art and include *Agrobacterium*-based techniques and techniques that do not require *Agrobacterium*. Non-*Agrobacterium* techniques involve the uptake of exogenous genetic material directly by protoplasts or cells. This can be accomplished by PEG or electroporation mediated uptake, particle bombardment-mediated delivery, or microinjection. In each case the transformed cells are regenerated to whole plants using standard techniques known in the art.

20 Methods for transformation of many dicot and monocot species are well-known in the art. Preferred techniques include direct gene transfer into protoplasts using PEG or electroporation techniques, particle bombardment into callus tissue, as well as *Agrobacterium*-mediated transformation.

In addition, the candidate variant library protein may also be made as a fusion protein, using techniques well known in the art. For example, the variant protein may be fused to other proteins to increase expression or stabilize the protein. Similarly, other fusion partners may be used, such as antibodies, targeting sequences that allow localization of the library members into a subcellular or extracellular compartment of the cell, rescue sequences or purification tags, that allow the purification or isolation of either the library protein or the nucleic acids encoding them; stability sequences, which confer stability or protection from degradation, fusion proteins including reporter, detection and selection genes or proteins, or combinations of these, as well as linker sequences as needed.

35 In a preferred embodiment, the candidate variant proteins or candidate variant library proteins are purified or isolated after expression. Variant proteins may be isolated or purified in a variety of ways known to those skilled in the art depending on what other components are present in the sample. Standard purification methods include electrophoretic, molecular, immunological and chromatographic techniques, including ion exchange, hydrophobic, affinity, and reverse-phase HPLC chromatography, and chromatofocusing. Ultrafiltration and diafiltration techniques, in conjunction with protein concentration are also useful. For general guidance in suitable purification techniques, see Scopes,

WO 02/090300

PCT/US02/14358

R., Protein Purification, Springer-Verlan, NY (1982). The degree of purification necessary will vary depending on the use of the variant protein. In some instances, no purification will be necessary.

Once made, the variant TR proteins may be experimentally tested and validated in in vivo and in vitro assays. Suitable assays include primary and secondary screening assays and characterization of purified protein kinetic parameters, i.e., K_{cat} and K_m (See Figures 11 and 12).

Once made, the variant TR proteins and nucleic acids of the invention find use in a number of applications. In a preferred embodiment, the variant TRs are used to reduce the antigenicity of glutens in wheat, rye and barley.

In other embodiments, the variant TRs are used to reduce the disulfide bonds in toxic proteins, such as those found in snake venom, bees, scorpions and the bacterial neurotoxins tetanus and botulinum.

In a preferred embodiment, the variant TRs are used to reduce alternative substrates. Alternative useful substrates for thioredoxin reductases include a number of plant and mammalian proteins found to contain thioredoxin domains. For example, protein disulfide isomerase (PDI) contains two regions that exhibit internal sequence homology to thioredoxin. PDI is a substrate for thioredoxin reductase. Protein disulfide isomerases have been identified from mammalian sources, such as bovine (Yamauchi et al., Biochem. Biophys. Res. Commun. 146:1485-1492, 1987), chicken (Parkkonen et al., Biochem. Zn 256:1005-1011, 1988), human (Rajilajaniemi et al. EMBO J. 6:643-649 1987), mouse (Gong, et al., Nucleic Acids Res. 16:1203, 1988), rabbit (Fliegel et al., J. Biol. Chem. 265:15496-15502, 1990), and rat (Edman et al., Nature 317:267-270, 1985). PDI has been isolated from yeast (Tachikawa et al., J. Biochem. 110:306-313). Suitable PDIs can be found in WO9501425 published 19950112 and WO9500636 published 19950105, as well as other PDIs known in the art including human and plant forms.

Compositions and uses of redox agents that are substrates of thioredoxin reductase, such as thioredoxin and PDI, are known in the art, and are discussed herein. Disulfide linkages are present in many types of proteins such as enzymes, structural proteins, etc. Enzymes are catalytic proteins such as proteases, amylases, etc., while structural proteins can be scleroproteins such as keratin, etc. Protein material in hair, wool, skin, leather, hides, food, fodder, is stains, and human tissue contains disulfide linkages. Treatment of some of these materials with PDI and thioredoxin, and a redox partner have been described previously. By way of example, the use of thioredoxin for waving, straightening, removing and softening of human and animal hair is described EP 183506 and WO8906122. US Patent 4771036 also describes the use of thioredoxin for prevention and reversal of cataracts. Use of thioredoxin to prevent metal catalysed oxidative damage in biological reactions is described by Pigiet et al. in EP 237189. EP 272781 and EP 276547 describe the use of PDI for reconfiguration of human hair, and for treatment of wool, respectively. The uses of such enzymes

WO 02/090300

PCT/US02/14358

have all been connected with reduction of protein disulfide linkages to free protein sulhydryl groups and/or the rearrangement of disulfide linkages in the same or between different polypeptides.

Consequently, thioredoxin reductases of the invention can be added to such compositions as a redox partner, optionally with its cofactor NADH or NADPH, to regenerate the redox agent and thus

- 5 enhance the compositions' usefulness. In an alternative embodiment, the thioredoxin variant of the invention are provided as protein fusions with the redox agent as taught herein. For example, the compositions can be used for the treatment or degradation of scleroproteins, especially hair, skin and wool, dehairing and softening of hides, treatment and cleaning of fabrics, as additives to detergents, thickening and gelation of food and fodder, strengthening of gluten in bakery or pastry products, and
- 10 as pharmaceuticals for the alleviation of eye sufferings. The compositions of the invention, particularly with PDI, can be used with other protein containing materials to generate intermolecular protein disulfide cross-links yielding high molecular weight or gelled compositions. Thus the present invention can be used in the field of food processing such as of raw fish meat paste, kamaboko (fish cake), fish/livestock meat sausage, tofu (soy bean curd), noodles, confectionery, bread, dough, food adhesives, sheet-like meat food, yogurt, jelly and cheese. In addition, they can also be used as novel
- 15 protein-derived materials in a wide range of industries including cosmetics, raw materials of microcapsules and carriers of immobilized enzymes.

In a preferred embodiment, variant TR-oleosin-thioredoxin and oleosin-variant thioredoxin-reductase fusion proteins accumulate in association with the oil bodies. In an alternate embodiment, oleosin-thioredoxin/variant thioredoxin-reductase hybrid fusion proteins accumulate in association with the oil

bodies. The oil bodies can be fractionated to achieve partial purification of the fusion proteins.

Purified oil bodies, with the associated fusion proteins, can be used as ingredients for testing of thioredoxin and thioredoxin-reductase activity and functional benefits in dermal (cosmetics) or food

- 25 use applications. Oil bodies have very suitable processing and formulation characteristics for cosmetic and food ingredients. Therefore, delivery of thioredoxin and/or thioredoxin-reductase as oleosin fusions associated with oil bodies simplifies processing and increases product stability.

In an alternate embodiment, a second purification step can be performed to purify thioredoxin or thioredoxin-reductase from the oil bodies. This leads to a highly purified preparation of the proteins that can be used as an ingredient for testing the activity of thioredoxin and thioredoxin-reductase, and for providing functional benefits in cosmetics or food uses. See also U.S. Patent Publication No. 2002/0037303; incorporated herein by reference.

- 35 In addition to other formulations and composition embodiments discussed herein, e.g., oil body embodiments, the compositions of the invention can contain soluble thioredoxin reductases and/or redox agents, and other ingredients known in the art as e.g. excipients, stabilizers, fillers, detergents, etc. The compositions can be formulated in any convenient form, e.g. as a powder, paste, liquid or in granular form. The enzyme(s) may be stabilized in a liquid by inclusion of enzyme stabilizers. Usually,

WO 02/090300

PCT/US02/14358

the pH of a solution of the composition will be 5-10 and in some instances 7.0-8.5. Often a sterile composition is preferred depending on the use.

- 5 Additionally, grain and grain-derived product performance in livestock feed are also affected by inter- and intramolecular disulphide bonding. Grain digestibility, nutrient availability, and the neutralization of anti-nutritive factors (e.g., protease, amylase inhibitors etc.) would be increased by reducing the extent of disulphide bonding (see WO 00/36126, filed 15 December 1999). Expression of transgenic thioredoxin reductase variants, optionally with thioredoxin, in corn and soybeans and the use of thioredoxin reductase in grain processing, e.g., wet milling, provides an alternative method for
- 10 reducing the disulfide bonds in seed proteins during or prior to industrial processing. The invention therefore provides grains with altered storage protein quality as well as grains that perform qualitatively differently from normal grain during industrial processing or animal digestion (both referred to subsequently as "processing"). This method of delivery of thioredoxin reductase, optionally with thioredoxin, eliminates the need to develop exogenous sources of thioredoxin and/or thioredoxin reductase for addition during processing. A second advantage to supplying thioredoxin and/or thioredoxin reductase via the grains is that physical disruption of seed integrity is not necessary to bring the enzyme in contact with the storage or matrix proteins of the seed prior to processing or as an extra processing step. The invention described herein is applicable to all grain crops, in particular corn, soybean, wheat, and barley, most particularly corn and soybean, especially
- 20 corn. Expression of transgenic thioredoxin reductase, optionally with thioredoxin, in grain is a means of altering the quality of the material (seeds) going into grain processing, altering the quality of the material derived from grain processing, maximizing yields of specific seed components during processing (increasing efficiency), changing processing methods, and creating new uses for seed-derived fractions or components from milling streams. The invention thus provides a plant which expresses a thioredoxin reductase variant, optionally with thioredoxin, preferably under control of an inducible promoter, for example either operatively linked to the inducible promoter or under control of transactivator-regulated promoter wherein the corresponding transactivator is under control of the inducible promoter or is expressed in a second plant such that the promoter is activated by hybridization with the second plant; wherein the TR is preferably thermostable or a eukaryotic
- 25 reductase; such plant also including seed therefor, which seed is optionally treated (e.g., primed or coated) and/or packaged, e.g. placed in a bag with instructions for use, and seed harvested therefrom, e.g., for use in a milling process as described above. The transgenic plant of the invention may optionally further comprise genes for enhanced production of NADPH or NADH.
- 30
- 35 The invention further provides a method for producing starch and/or protein comprising extracting starch or protein from seed harvested from a plant as described above; and a method for wet milling comprising steeping seed from a thioredoxin reductase- expressing plant as described above and extracting starch and/or protein therefrom. Heat stable enzymes are preferred, such as from a

WO 02/090300

PCT/US02/14358

thermophilic organism, e.g., from an archaea, for example from *Methanococcus jannaschii* or *Archaeoglobus fulgidus*, e.g., as described herein.

Expression of transgenic thioredoxin reductase variants, optionally with thioredoxin, in grain is also useful to improve grain characteristics associated with digestibility, particularly in animal feeds.

- 5 Susceptibility of feed proteins to proteases is a function of time and of protein conformation. Kernel cracking is often used in feed formulation as is steam flaking. Both of these processes are designed to aid kernel digestibility. Softer kernels whose integrity can be disrupted more easily in animal stomachs are desirable. Conformational constraints and crosslinks between proteins are major determinants of protease susceptibility. Modifying these bonds by increased thioredoxin and/or
- 10 thioredoxin reductase expression thereby aids digestion. Protein content and quality are important determinants in flaking grit production and in masa production. Reduction of disulphide bonds alters the nature of corn flour such that it is suitable for use as a wheat substitute, especially flours made from high-protein white corn varieties. Over half of the US soybean crop is crushed or milled, and the protein quality in the resulting low-fat soy flour or de-fatted soy flour (or soybean meal) is important for subsequent processing. Protein yield and quality from soybean processing streams are economically
- 15 important, and are largely dependent upon protein conformation. Increasing thioredoxin activity through expression of transgenic thioredoxin and/or thioredoxin reductase increases protein solubility, and thus increases yield, in the water-soluble protein fractions. Recovery is facilitated by aqueous extraction of de-fatted soybean meal under basic conditions. Enhancing thioredoxin activity through
- 20 expression of transgenic thioredoxin and/or thioredoxin reductase also reduces the required pH for efficient extraction and thereby reduces calcium or sodium hydroxide inputs, as well as lowering the acid input for subsequent acid precipitation, allowing efficient recovery of proteins without alkali damage, and reducing water consumption and processing plant waste effluents (that contain substantial biological oxygen demand loads). Protein redox status affects important functional
- 25 properties supplied by soy proteins, such as solubility, water absorption, viscosity, cohesion/adhesion, gelation and elasticity. Fiber removal during soy protein concentrate production and soy protein isolate hydrolysis by proteases is enhanced by increasing thioredoxin activity as described herein. Similarly, as described for corn above, increasing thioredoxin activity through expression of transgenic thioredoxin and/or thioredoxin reductase enhances the functionality of enzyme-active soy flours and
- 30 the digestibility of the soybean meal fraction and steam flaking fraction in animal feeds. Modification of protein quality during seed development and during processing are both provided, although it is preferred that the transgenic thioredoxin and/or thioredoxin reductase be targeted to a cell compartment and be thermostable, as described above, to avoid significant adverse effects on storage protein accumulation possibly encountered as a result of thioredoxin activity during seed development. Alternately, the thioredoxin reductase variant, and optionally thioredoxin, can be added
- 35 as a processing enzyme, (or as fusions as taught herein) as (in contrast to corn wet milling) breaking the disulphide bonds is not necessary until after grain integrity is destroyed (crushing and oil extraction). Protein disulfide isomerase (PDI) are also useful as described above for thioredoxin.

WO 02/090300

PCT/US02/14358

Regarding use of oil bodies with TR, incorporated herein by reference is US20020037303 entitled "Thioredoxin and thioredoxin reductase containing oil body based products" published 20020328.

- Additional uses of the enzymes of the invention for seed and grain can be found in WO0058453, published 20001005. Thioredoxin reductase variants can be expressed optionally with thioredoxin, or added exogenously, for the uses described therein for seed and grain quality enhancement. The transgenic plant of interest include is barley, wheat, Arabidopsis, tobacco, rice, Brassica, Picea, or soy bean, maize, oat, rye, sorghum, millet, triticale, and forage and turf grass. A transgenic plant of the invention can have reduced allergenicity in comparison to the same part of a non-transgenic plant of the same species. The allergenicity can be hypersensitivity, wherein said hypersensitivity is reduced by at least 5%. Further, a transgenic plant of the invention can have increased digestibility in comparison to the same part of a non-transgenic plant of the same species. The digestibility is increased by at least 5 percent. A transgenic plant can have at least part of said plant having an earlier onset and/or an increased expression of a gibberellic acid inducible enzyme in comparison to the same part of a non-transgenic plant of the same species. Preferably the enzyme is pullulanase, alpha-amylase. The parts of the plant are preferably edible parts, more preferably grain or seed. Preferred promoters are a seed or grain maturation-specific promoter, e.g., selected from the group consisting of rice glutelins, rice oryzins, rice prolamines, barley hordeins, wheat gliadins, wheat glutelins, maize zeins, maize glutelins, oat glutelins, sorghum kasirins, millet pennisetins, rye secalins, and a maize embryo-specific globulin. In other embodiments are a food, feed or beverage product made from the transgenic seed or grain of the invention. The food, feed, or beverage can be flour, dough, bread, pasta, cookies, cake, thickener, beer, malted beverage, or a food additive. The food, feed, or beer product of can have reduced allergenicity and/or increased digestibility. Further, a dough product can have increased strength and volume in comparison to a dough made from a non-transgenic seed or grain of the same species. The food, feed, or beverage can have hyperdigestible protein and/or hyperdigestible starch. The food, feed, or beverage can be hypoallergenic. The above embodiments are also achieved by exogenous addition of the enzymes of the invention, as would be known in the art. It has been shown that reduction of disulfide protein allergens in wheat and milk by thioredoxin decreases their allergenicity. Thioredoxin treatment also increases the digestibility of the major allergen of milk (beta-lactoglobulin), as well as other disulfide proteins. A more detailed discussion of the benefits of adding exogenous thioredoxin to food products is presented in U.S. Patent No. 5,792,506, which is specifically incorporated herein by reference. The compositions and methods can be enhanced using the TR variants of the invention.
- As discussed herein, the proteins of the invention can be used to reduce allergenicity of proteins in food and feed. For example, see US6190723 and reference therein, which is specifically incorporated herein by reference, for uses of thioredoxin with thioredoxin reductase and NADPH as exogenously added treatments. Skin tests and feeding experiments carried out with sensitized dogs showed that treatment of their food prior to ingestion eliminated or decreased the allergenicity of the food.

WO 02/090300

PCT/US02/14358

Consequently, provided herein are compositions for and methods of decreasing the allergenicity of an allergenic food or feed protein. The food or feed protein or food or feed containing the protein or proteins is contacted with an amount of thioredoxin, thioredoxin reductase, and cofactor, namely NADPH, NADH or combination thereof, effective for decreasing the allergenicity of the protein. This can be followed by administering the contacted protein to an animal or human, wherein the allergic symptoms exhibited by the animal or human are decreased as compared to a control. The allergenic food/feed protein is preferably from the beef, cow's milk, egg, soy, rice and wheat proteins. Also embodied are ingestible food/feed products containing thioredoxin and TR variant and further containing cofactor. The enzymes made be exogenously added, or one or the other may be transgenically or naturally present, singly or as a fusion. The ingestible food is preferably hypoallergenic because of the treatment. The food product can be a pet food or baby food or formula. The food product can contain beef, egg, soy, wheat or milk protein. It can be an ingestible meat food product. US Patent 5792506 is and its references are incorporated by reference.

Similarly, in US6114504 compositions and methods of reducing cystine containing animal and plant proteins, and improving dough and baked goods' characteristics is provided which includes the steps of mixing dough ingredients with a thiol redox protein to form a dough and baking the dough to form a baked good. The method of the present invention preferably uses reduced thioredoxin with wheat flour which imparts a stronger dough and higher loaf volumes. The methods and compositions are enhanced using the proteins of the invention. A method of reducing a glutenin or gliadin protein is by adding thioredoxin to a liquid or substance containing said glutenin or gliadin protein; reducing the thioredoxin by means of thioredoxin reductase variant and a cofactor, namely NADPH, NADH or combination thereof, and reducing the glutenin or gliadin protein by the reduced thioredoxin. A composition contains a glutenin or gliadin protein, added or endogenous thioredoxin, added or endogenous (as from a transgenic plant) thioredoxin reductase variant, and added cofactor, namely NADPH, NADH or combination thereof. The method is useful to reduce any water insoluble or soluble, seed-derived protein comprising. One can add thioredoxin to a liquid or substance containing said protein; reducing the thioredoxin by means of thioredoxin reductase variant and its cofactor, namely NADPH, NADH or combination thereof.

The invention is also useful for increasing hyperdigestibility of food and feed proteins. See US Patent 5952034 that provides for compositions and methods to increase the digestibility of food proteins by thioredoxin reduction. The methods are enhanced by use of the enzymes of the invention.

Compositions and method of increasing the digestibility of a food comprise treating a food with an amount of thioredoxin, thioredoxin reductase variant, and its cofactor, namely NADPH, NADH or combination thereof, effective for increasing the digestibility of the food; and optionally administering the treated food to an animal or human thereby increasing the digestibility of the food as measured by the symptoms exhibited by said animal or human as compared to a control. The food preferably

WO 02/090300

PCT/US02/14358

contains milk or wheat or eggs. In the above embodiments, the thioredoxin reductase variant can be provided as a protein fusion with thioredoxin.

The compositions of the invention also find additional uses. Thioredoxin and other redox agents, such as PDI, are known to be useful in protection against stress and injury. Accordingly, the compositions of the invention can be used to enhance redox agent compositions for such treatment. In one embodiment, TR variants are used to manipulate nitrosative stress to upregulate nitrosative stress defenses. See US6359004. Thioredoxin can act as a radical scavenger, thus disease and conditions related to free radicals can be treated with TR variants, preferably in combination with thioredoxin. Thus, in one aspect, the present invention provides compositions and methods for the prevention or treatment of eye diseases, such as cataracts. In another aspect, the present invention relates to the prevention or treatment of diseases caused by oxidative stress or having oxidative stress as a component. See for example US patent 6379664. In one embodiment is provided compositions and methods of inhibiting or reversing the formation of a cataract in an eye, by contacting the eye with an effective cataract-inhibiting amount of a composition of the invention, containing TR variant, preferably in combination with thioredoxin. In another embodiment, intraocular injection of thioredoxin in combination of a TR variant and cofactor suppresses retinal photooxidative stress, and as a therapeutic strategy to prevent retinal photic injury. In another embodiment, compositions of the invention containing thioredoxin activity are useful to treat or minimize oxidative stress and ischemia-reperfusion induced in acute lung injury. And consequently further finds use in lung transplantation, particularly in patients with end-stage lung diseases, such as cystic fibrosis, emphysema, pulmonary fibrosis, and pulmonary hypertension. The compositions of the invention find use as storage compositions to maintain integrity of organs for transplant. In another embodiment, thioredoxin in combination with the TR variants promotes the in vitro survival of primary cultured neurons. Further the compositions will provide a neuroprotective effect in the penumbra to modify neuronal damage during focal brain ischemia. The compositions will also provide protection and improvement of motoneurons from or after nerve injury. In another embodiment, compositions of the invention protect the retina from ischemia-reperfusion injury. Burn injuries can also be treated with compositions of the invention. Thioredoxin and TR variants provide a rapid antioxidant defense, improves coagulation processes, cell growth, and control of the extracellular peroxide tone intimately linked to cytoprotection and wound healing in burns. Finally, the compositions of the invention provide thiol-antioxidants that are good candidates for controlling Epstein-Barr virus (EBV) infection.

TR variants can provide direct benefit by removing deleterious ascorbyl free radical and dehydroascorbate, which are reduced to ascorbic acid by thioredoxin reductase. Thus TR provides a direct antioxidant effect and treatment. The compositions can optionally contain cofactors.

In the diseases and conditions described herein, the TR variants can be supplied alone or in combination with thioredoxin or other redox agents and cofactors. The enzymes by be separate or fused. The TR variant may act with host redox agents or redox agent can be exogenously added.

WO 02/090300

PCT/US02/14358

The following examples serve to more fully describe the manner of using the above-described invention, as well as to set forth the best modes contemplated for carrying out various aspects of the invention. It is understood that these examples in no way serve to limit the true scope of this invention, but rather are presented for illustrative purposes. All references cited herein, including U.S.S.N. 60/289,029, filed May 4, 2001, U.S.S.N. 60/370,609, filed April 5, 2002, and the provisional application by Desjarlais and Muchhal, entitled "Novel Nucleic Acids and Proteins with Thioredoxin Reductase Activity", filed April 29, 2002, serial number not assigned, are incorporated by reference.

EXAMPLES

Example 1

Computational Design of Variant Proteins

Overview

The initial PDA™ design strategy for creating variants with improved NADH-dependent TR activity is detailed below. In short, the structural information from both E coli and Arabidopsis enzymes, and the co-factor conformation diversity was used to design two different libraries (referred to as TR-1 and TR-2 henceforth), each with ~2000 combinatorial members.

Wildtype TR genes used as scaffold proteins:

- 1) Arabidopsis NTR1 gene cloned in pET29a. The encoded protein has an N-terminal S-tag. The protein may be expressed using BL21-S1 cells (salt induced) or BL21-Star (IPTG induced), lysed using BugBuster HT.
- 2) Thioredoxin j. A codon-optimized gene synthesized and cloned in pDEST-14, expressed in BL21-S1-Star. Soluble fraction used as substrate during primary screenings. N- and/or C-terminal His tagged versions made. The C-terminal His-tagged TRx purified by affinity chromatograph for use in kinetic determinations.

Assay: Kinetic assay based on continuous detection of formation reduced product of DTNB at 412 nm.

A more detailed overview of the screening strategy used for identification and kinetic characterization of "hits" is described in Figure 4.

Purified proteins were used for all the kinetic characterizations and second and third tier screenings. High throughput procedures for generating required amounts of purified proteins were either independently developed or adapted from existing commercial protocols. A snapshot of these methods is presented in Figure 5. The detailed protocols used for high-throughput culture, induction, expression, protein purification and enzymatic characterization are described below.

The kinetic parameters (Km and Kcat) for the purified WT NTR-1 enzyme (unmodified) with respect to

WO 02/090300

PCT/US02/14358

both the NADH and NADPH substrates to define the benchmark for PDA™ designed variants. The WT enzyme has ~ 4 fold higher Kcat (equivalent to the Vmax using 1 ug of TR protein) for the native (NADPH) co-factor than NADH. Also the Km is ~ 50 fold higher for NADH compared to NADPH. The data for WT enzyme is presented in Figure 6.

5

The TR Libraries were constructed using standard molecular biology procedures of site-directed mutagenesis and recursive PCR. Combinatorial pieces representing specifically mutated gene segments were joined together using specific restriction enzymes. The quality of these libraries was evaluated from sequence and expression analysis of randomly picked clones. These details for the

10

TR-1 and TR-2 are presented in Figures 7 and 8 respectively. In addition to these combinatorial libraries, individual C-region combinations for each of these two libraries (24 for TR-1 and 48 for TR-2) were synthesized in WT backbone to evaluate the effect of this critical region identified by PDA™, these clones are henceforth referred to as "defined clones" along with the individual members of TR-3 and TR-4 (see below).

15

A computationally relevant description of the two libraries is presented in Figures 9A and B. The designed positions (orange) and the docked co-factor (blue or yellow) with appropriate conformation are identified.

20

In addition to these two libraries, a couple of very small libraries were generated to explore additional strategies. TR-3 had 18 members and was designed as a fine tuning approach based on results for the best clone from TR-2 screening. TR-4 had 16 members and was based on sequence alignment of TR and AhpF sequence. AhpF codes for a NADH dependent peroxiredoxin reductase, an activity analogous to TR.

25

The summary of results from the screening of these 4 libraries is presented in Figure 10.

The screening of TR-1 library did not identify any clones with significantly improved TR activity with NADH as a co-factor, compared to WT NTR-1. This likely the result of using the "incorrect" co-factor conformation.

30

The TR-2 library had several clones with significantly improved NADH-dependent activities. Two of the best variants with different C-regions sequences were "RYN" and "RFN". Mutations in other designed positions did not have a significant effect on the overall properties of the TR enzymes. The following slides present detailed kinetic data for many of these variants.

35

M-RYN, L-RYN and WT kinetic parameters and their activities at different co-factor concentrations are described Figures 11A and B respectively. Both of these variants have significantly higher NADH-dependent activities compared to WT. In addition they have significantly reduced NADPH dependent

WO 02/090300

PCT/US02/14358

activity. This is termed "Co-factor Switch". At co-factor concentrations of 2.5 mM and above both of these PDA™ designed NTRs have >50% of WT NADPH activity with NADH as co-factor.

5 The sequence alignment of these clones and their relative computational ranking from the design perspective is shown in Figure 17A.

10 The presence of N in RYN and RFN clones created a potential glycosylation site. This site was "designed out" using PDA™ without affecting the activity profile of these clones significantly. The data and strategy for this is described below.

Computational representation of the critical RRR to RYN change is described in Figure 18.

15 In addition to RYN and RFN combinations in the C-region, REN, RLN, RRN combinations also had significantly improved NADH-dependent activity. The RRN variant also maintained its WT level of NADPH dependent activity. This data is summarized in Figure 12. Additionally, RRT, RYT, RLR, KYN, MYN, QYN C-region variants also showed improved NADH-dependent activity.

20 The results from screening of these libraries point strongly to the significance of three RRR residues in the C-region for determining the co-factor specificity profile. To address the significance of all possible combinations of 20 amino acids at each of these positions, a high complexity random RRR library was designed and screened to identify the best variants for their activity with NADH. An oligonucleotide with NNK degeneracy at each of the three R positions was used to construct this library with a theoretical combinatorial potential of 32768 members.

25 After screening only a small proportion of this library, the sequence and activity analysis of the best clones indicated that a R to W mutation at the first R position had the most interesting activity profile. This is also substantiated from the bioinformatics analysis of most naturally occurring NAD(P)H dependent enzymes sequences suggesting the presence of an aromatic amino acid. This led us to design a PDA™ library where the first R is forced to be an aromatic amino acid during PDA™ simulations. This led to the design of two additional smaller PDA™ libraries called R1-W and WXX. 30 The computational strategy for their design is described below.

35 The best hits from all these new library designs were analyzed (using purified enzymes) for their relative activity at 0.6 and 1.2 mM each of the two co-factors. Their Km and Kcats were also determined and the data is presented in Figures 13 A and B respectively. These clones have "highly improved" NADH dependent TR activities. In addition to their improved NADH activity, some of the variants also have improved NADPH dependent activities. This in essence represents creating TR variants with better catalytic efficiencies for both the co-factors. This is also reflected in the several fold higher NADH Kcat values for all the variants. The Km for NADH

WO 02/090300

PCT/US02/14358

remained unchanged for most of the improved variants, except WRT which has a two fold reduced K_m for this co-factor. The members of this list coming from either R1-W and WXX libraries are indicated in Figure 13C. A computational model of the two best clones from R1-W library are depicted in Figure 14 for a structural perspective on their activity.

5

The PDA™ Design process for TR has thus identified :

- Five or more variants with equal to or better than 50% of WT NADPH activity, with NADH at 1.2 mM.
- At least one variant meets this activity milestone even at 0.6 mM NADH
- A large number of these variant have improved catalytic efficiency for the NADPH activity also.
- The best variant has a 13-fold better K_{cat}/K_m and 2-fold lower K_m for NADH compared to WT

10

15 Thioredoxin Reductase R1-W Library

A new set of PDA™ simulations was performed to evaluate the use of an aromatic amino acid (F, Y, or W) at the first position of the trio of residues discovered by Xencor to be extremely important in modulating activity levels with NADH and NADPH (corresponding to the position of R in the RYN variants). The new simulations were motivated by the observation that a small number of NAD(P)H utilizing enzymes contain an aromatic at this position, and the potential for a stacking interaction between the aromatic and the adenine ring on NAD(P)H.

20

Simulation of 20^{10} (10^{15}) sequences resulted in the library shown below, which defines 1296 variants for in vitro screening. The 10 positions were selected by structural analysis of critical residues for cofactor binding. Analysis of the simulation results revealed that sampling amino acid diversity at 6 of the 10 positions would result in a high-quality library of modest size.

25

The 4th PDA™ library, with diversity at 6 positions, in the context of W versus R at one position, is defined as:

30

L I R R R V I (wt)

L I W R T V I

A L A S I V

F V C N

35

E C

K

L

M

Q

WO 02/090300

PCT/US02/14358

S

High throughput screening of this library yielded the following high activity WXX clones. These clones have been ranked computationally by performing PDA™ simulations that represent the 4th PDA™ combinatorial library.

Out of the 1296 possible sequences in this library the highly active WXX clones rank computationally as follows:

LIWRTVI	13/1296	(rank/library size)
LIWLSVI	51/1296	
LIWMSVI	26/1296	
LIWRSVI	46/1296	

Note that these rankings are not intended to be predictive of relative activity: the calculation was designed to define the broadest set of structurally compatible cofactor binding pocket diversity in the smallest number of sequences. All of the library members are in the top 0.001% of the 20⁶ theoretically possible sequence combinations at the 6 positions included in the 4th library, demonstrating a focusing effect of over 10⁴. This furthermore constitutes a focusing effect of at least 10⁹ relative to the 20¹⁰ sequence combinations included in the original simulation.

Note also that these rankings are based purely on simulated interaction with NADH. They do not take into account the specificity of the enzyme for or against NADPH. Since the project objectives did not include NADPH/NADH specificity, comparative modeling of the two cofactor-protein complexes was not performed.

Additional Variants

Based on the success of the R1-W library, and the observation of considerable diversity at the 2nd and 3rd R positions in both the simulations and laboratory screening, Xencor constructed a small complexity (400) library to sample all possible WXX combinations. High throughput screening of this library led to the discovery of several additional variants with high activity using NADH, and variable activity using NADPH.

The 5 best clones from this library, containing diversity only at the 3 RRR positions, are listed below. While the design of this library was directly influenced by all of the previous PDA™ simulation and experimental results, the library was not based on a PDA™ simulation per se. Thus there are no computational rankings for these variants.

WIS

WFQ

WVR

WO 02/090300

PCT/US02/14358

WMG

WVG

Computational Rankings of RYN Thioredoxin Reductase Variants

- 5 The individual "RYN" clones have been ranked computationally by performing PDA™ simulations that represent the 2nd PDA™ combinatorial library constructed and screened by Xencor. Simulation of 20⁸ (2.5 x 10¹⁶) sequences resulted in the library below, which defines 2304 variants for in vitro screening. The 8 positions were selected by structural analysis of critical residues for cofactor binding.

10

The 2nd PDA™ library, with diversity at 8 positions is defined as:

L I G D R R R S

Q M S N K Y T D

- 15 L Q E N

L I

Out of the 2304 possible sequences in this library the wild-type and highly active RYN clones rank as follows:

20	LIGDRRRS (wt)	329
	LIGDRYNS	339
	LLGDRYNS	698
	LMGDRYNS	920

- 25 Note that the rankings are not intended to be predictive of relative activity: the calculation was designed to define the broadest set of structurally compatible cofactor binding pocket diversity in the smallest number of sequences. All of the library members are in the top 0.00001% of the 20⁸ theoretically possible sequence combinations at the eight positions included in the 2nd library, demonstrating a focusing effect of over 10⁷.

30

Note also that these rankings are based purely on simulated interaction with NADH. They do not take into account the specificity of the enzyme for or against NADPH. Since the project objectives did not include NADPH/NADH specificity, comparative modeling of the two cofactor-protein complexes was not performed.

35

Novel Thioredoxin Reductase Variants

Low Complexity Library. The initial success of the RYN variant motivated Xencor to pursue further optimization of this variant by refining the amino acids in the RYN variant, leading to the very small 18-member library shown below.

WO 02/090300

PCT/US02/14358

R R R
M Y N
F D

5

Screening of this library revealed that the **RFN** combination was of similar activity to the RYN variants discovered previously. According to PDA™ simulations, this clone ranks 7th in this library (RYN ranks 3rd).

- 10 **Non-glycosylation variants.** Because of the inadvertent introduction of a potential N-linked glycosylation site (consensus N-X-[T/S]) in the RYN and related variants (**RYDAFNASKIMQQ**), PDA™ simulations were performed to assess the feasibility of extinguishing the potential site by substitution of the Serine (S) two positions downstream of the Asn (N) in the RYN variants. The simulations indicate that several amino acid substitutions would be favorable, including Ser to Ala, which Xencor then produced and characterized experimentally. In this one-position simulation (NAX),
- 15 Ala ranked 6th, with Thr and Ser ranked 1st and 2nd, respectively. Experimental data indicates that the Ala substitution has no detectable effect on the activity of the RYN variants.

RYN-A (339/2304, 6/20) (rank/original library size, rank/NAX library size)

20 **RFN-A** (7/18, 6/20)

Computational Strategy

Primary Goal: Conversion of arabidopsis thioredoxin reductase activity such that it efficiently utilizes NADH vs. NADPH

25

Basic Outline of Strategy:

I. generate starting model

- 30 use E coli structure (1TDF) to "graft" coordinates of NADP cofactor into coordinate frame of arabidopsis structure (1VDC), which does not include cofactor coordinates.

II. define working cofactor conformation

- a. direct derivation by deleting P from NADP
- 35 b. indirect derivation by superposition of NAD coordinates from various NAD-utilizing enzymes

III. run PDA simulation(s) to generate combinatorial library possibilities.

- a. define library positions
- b. run simulation(s)

WO 02/090300

PCT/US02/14358

c. generate library

Detailed Outline of Strategy

I. Generation of starting model

5

A. The 1VDC structure file was processed to create a more reasonable numbering system for the structure (the original version contained an atypical numbering format so that the numbering agreed with the E coli structure).

10 B. Structure alignment for grafting NADP coordinates from 1TDF to 1VDC

An alignment was obtained using the C-alphas from the following residues: 117, 119, 151-156, 174-181, and 242-244. This gives an RMSD of 0.48 Å for 19 matched atoms (with a maximum deviation of 0.89 Å).

15 C. Note that no minimization was done on the final model.

II. Defining the working cofactor conformations

A. The initial cofactor conformation was defined simply by deleting the phosphate group from the NADP cofactor contained within the 1TDF file. We will refer to this conformation as NAD_TDF.

20

B. Alternative NAD conformations.

Adam Thomason developed Perl scripts that scan the PDB for structures containing NAD cofactors. The scripts then perform a full or partial superposition of the NAD from the extracted PDB file onto the reference NAD_TDF. A large number of NAD conformations were thus collected (see Fig. 19) and ready for use in PDA simulations.

25

Simulations have been performed using either the NAD_TDF conformer or the NAD_GRB conformer (from 1GRB - human glutathione reductase), which had the lowest all-atom r.m.s.d to NAD_TDF.

Visual inspection of over 100 NAD conformers indicates that the ribose pucker found in NAD_GRB is significantly more prevalent than that in NAD_TDF, suggesting that this conformer is of lower energy. It is possible that the rare conformer seen in NAD_TDF stems from the fact that this conformer was derived from NADP coordinates.

30

C. Hydroxyl rotamer states.

35 The orientation of the hydrogen of a hydroxyl group can have a significant influence on side chain-cofactor interactions, particularly with respect to hydrogen bonding interactions. For library 1, a static pair of hydroxyl rotamers was utilized, because only a single ligand state can be included per simulation within the Xencor implementation of PDA™. Subsequently, the SPA package was developed such that a combinatorial set of ligand states can be included in the simulation. A support

WO 02/090300

PCT/US02/14358

program named "makeligands" (from makeligands.f90) was also developed to generate combinatorial sets of hydroxyl rotamer orientations.

III. PDA simulation(s) to generate combinatorial libraries

5

A. Defining library positions

The current strategy is to enhance interactions between the TRR protein and the adenine portion of NADH, particularly with the diol group on the adenine ribose, which is left behind when the phosphate is removed (see Fig. 20).

10

B. Library 1 Calculations - performed with PDA™

The first combinatorial library was generated using the PDA™ simulation package. In this package, ligands are incorporated as part of the "template", which restricts the number of ligand states per simulation to 1. Therefore, the hydroxyl rotamers on the adenine diol were arbitrary for this set of calculations. Furthermore, no charges were created for the NAD. The first set of calculations included several amino acid possibilities at position 189. For all subsequent calculations, the identity at this position was restricted to Histidine.

15

C. Library 1 definition

The rationale for library 1 was based on a combination of (i) quality of residues as predicted by ORBIT (based on probability tables generated by an ORBIT monte carlo simulation); (ii) structural intuition; and (iii) an emphasis on sampling a diversity of amino acid properties. At all positions, the wild type residue was included in the library. The most intriguing aspects of the library are various potential hydrogen-bonding interactions between side chains and the cofactor, giving rise to residues EDT at position 127, QE at position 195, EQ at position 217, and E at position 255. Because most NADH-utilizing enzymes contain an interaction between a carboxylic side chain and the adenine diol, the prediction of Q and E at position 195 is encouraging.

20

25

TRR Library 1:

30

127 LEDTA

165 IML

166 G

167 G

189 H

35

190 RYM

191 RQ

195 RYQE

217 SEQ

255 IE

WO 02/090300

PCT/US02/14358

D. Library 2 calculations - performed with SPA

Several simulations, using various cofactor conformations and sampling strategies, were performed for the development of library 2.

5

(i) The first set of simulations was performed using the NAD_TDF cofactor conformation for the heavy atom coordinates. Using this conformation, and 36 (6 x 6) hydroxyl rotamer combinations on the adenine diol, simulations were performed with either backbone ensemble or sub-rotamer sampling strategies.

10

(ii) The second set of simulations was performed using the NAD_GRB cofactor conformation for the heavy atom coordinates. Using this conformation, and 36 (6 x 6) hydroxyl rotamer combinations on the adenine diol, simulations were performed with either backbone ensemble or sub-rotamer sampling strategies.

15

E. Library 2 definition

The rationale for library 2 was based on a combination of (i) quality of residues as predicted by SPA (based on output free energy matrices and comparison of matrices from different simulations); (ii) structural intuition; (iii) an emphasis on sampling a diversity of amino acid properties; and (iv) feedback from Library 1 screens. At all positions, the wild type residue was included in the library. As before, the most intriguing aspects of the library are various potential hydrogen-bonding interactions between side chains and the cofactor. However, because an alternative cofactor conformer was used in these calculations, new sets of interactions are predicted by SPA, giving rise to residues Q at position 127, S at position 167, TN at position 195 (Fig. 3 A,B), D at position 217, and E at position 255. The S167 (Fig. 3C) was chosen despite a high free energy value, based on its predicted ability to hydrogen bond to the AO2* oxygen of the adenine diol and the supposition that a small movement would relieve the van der Waals clash. An additional residue N at position 169 was added to this library, based on the possibility that neutralizing the negative charge at this position would assist in improving binding affinity of the cofactor (note that N is a conservative mutation as it is found in the E coli TRR).

30

Most of the residues in library 2 were chosen based on simulations with NAD_GRB. However, I195 was added based on a high propensity for this residue in SPA calculations using the NAD_TDF cofactor conformation.

35

TRR Library 2:

126 123 118 R

1

WO 02/090300

PCT/US02/14358

127	124	119	L	Q	2		
128	125	120	S		1		
164	161	150	V		1		
165	162	151	I	M	L	3	
166	163	152	G			1	
167	164	153	G	S		2	
168	165	154	G			1	
169	166	155	D	N		2	
170	167	156	S			1	
189	186	175	H			1	
190	187	176	R	K	Q	3	
191	188	177	R	Y	E	L	4
192	189	178	D				1
193	190	179	A				1
194	191	180	F				1
195	192	181	R	T	N	I	4
196	193	182	A				1
216	213	202	S				1
217	214	203	S	D			2
218	215	204	V				1
254	251	242	A				1
255	252	243	I				1
256	253	244	G				1
							2304

ASSAYSEXPRESSION

- 5
1. The NTR coding region cloned in pET29 is expressed in BL21 Star (Invitrogen) cells. The volumes described here are typical for getting > 50 ug of purified protein, and can be either scaled up or down based on requirements.
 2. Inoculate colonies in a 96-deep well plate containing 1.5ml CG + Kanamycin (100ug/ml), inoculate appropriate controls. Grow overnight cultures at 37°C, 250rpm
 3. Next day, inoculate 200μl of overnight cultures in 5ml CG + Kanamycin (100ug/ml) in 4x24-well plate for each 96 deep well plate. Grow at 30°C, 250rpm, for 3hrs
 4. Make glycerol stocks from remaining overnight cultures and freeze at -80°C
 5. Induce the 5ml cultures with 1M IPTG to final concentration of 1mM. Grow overnight at 30°C, 250rpm
- 10
- 15

WO 02/090300

PCT/US02/14358

6. Next day, spin down the cells at maximum speed (Avanti J-20, 5300rpm) for 10min. Discard supernatant, pellets can be frozen at -80°C or proceed to S.tag Purification procedures

5 S.TAG PURIFICATION FOR 96-WELL PLATE

(96 samples (from cell pellets; Novagen, cat# 69232-3)

The S.Tag Thrombin Purification Kit uses a unique strategy that employs Biotinylated Thrombin, which enables simple and specific removal of the enzyme after digestion with Streptavidin Agarose.

- 10 The standard protocol calls for batch-wise binding to S-protein Agarose, washing, treatment with Biotinylated Thrombin, and capture with Streptavidin Agarose, leaving the purified protein in solution.

Kit Components

Components	Provided Volume	Vol for 1kit/24samples
S-protein Agarose (50% slurry in 50mM Tris-HCl, pH7.5, 150mM NaCl, 1mM EDTA, 0.02% sodium azide)	2ml	167µl slurry/sample
10X Bind/Wash Buffer (200mM Tris-HCl pH7.5, 1.5M NaCl, 1% Triton X-100)	3x5ml	100ml of 1X 1ml/sample
10X Thrombin Cleavage Buffer (200mM Tris-HCl pH8.4, 1.5M NaCl, 25mM CaCl ₂)	3ml	30ml of 1X 400µl/sample
Biotinylated Thrombin	50U (1.5U/µl)	25U (16.6µl) 1U (0.66µl)/sample
Streptavidin Agarose (50% slurry in phosphate buffer, pH7.5, 0.02% sodium azide)	2x0.4ml	1.6ml slurry 60µl slurry/sample

- 15 Additional materials:

Whatman Unifilter, 96-well, 800µl (Fisher, cat# PF7700-2804)

Bug Buster Protein Extraction Reagent (VWR, cat# 80500-208)

Protocol (5ml expression cultures)

- 20
1. Thaw frozen pellets (5ml) at RT for ~30min
 2. Add 500µl of Bug Buster HT, vortex to resuspend pellets and shake at RT for 20min
 3. Spin at max speed or 3000xg for 20min. Transfer supernatant (cell lysate) containing soluble proteins to a new plate.
 4. Use 150µl of cell lysate for purification, save remainder for later use
- 25 For 150µl

WO 02/090300

PCT/US02/14358

Adjust Tris-HCl and NaCl concentration to 20mM Tris and 150mM NaCl, pH7.5

	150μl	Bug Buster	$\times 100$
	10μl	1M Tris-HCl (final 20mM)	1ml
	15μl	5M NaCl (final 0.15M)	1.5ml
5	325μl	H ₂ O	32.5ml
	500μl	total	aliquot 350μl mix

5. Seal filter plate bottom with aluminum tape
6. Add 167μl of S-protein agarose mix using wide mouth tips
- 10 7. Add lysate (adjusted) to filter plate, seal plate with aluminum tape
8. Bind at RT for 30min – 1hr on an orbital shaker (Place plate on the side - Do not shake vigorously as this will tend to denature protein)
9. Remove aluminum tape from the bottom, apply vacuum
10. Wash 2 times with 500μl of 1X Bind/Wash Buffer, apply vacuum
- 15 11. Equilibrate 2 times with 1X Thrombin Cleavage Buffer with ~1X slurry volume = 200μl, apply very low vacuum
12. Re-seal filter plate bottom with aluminum foil
13. Make a mix of 1X Thrombin Cleavage Buffer and Biotinylated Thrombin

Master mix

Reagents	1Kit for 24 samples	
	each	X100
1X Thrombin Cleavage Buffer	80μl	8ml
Biotinylated Thrombin (1.5U/μl)	0.66μl	66μl
Aliquot		80.7μl

- 20 14. Gently shake tubes at RT for 1-2hr on micromixer setting = 5, amplitude = 4
15. Add 60μl slurry of Streptavidin Agarose
16. Incubate on orbital shaker at RT for 10min
17. Remove foil seal from the bottom of the filter plate
- 25 18. Spin at 500xg, 2min
19. To elute more protein, add 80μl of 1X cleavage buffer, spin at 500xg, 2min
20. Add equal volume of 50% glycerol, mix really well and store at 4°C temporary, for long-term storage, freeze at -80°C

30

BCA ASSAY

BCA Protein Assay Reagent Kit (Pierce, cat# 23227)

WO 02/090300

PCT/US02/14358

1. Preparation of standards and working reagent
 - a. Standards (working range is 0.125 - 2 µg/µl)

Tube	Vol of Diluent (µl)	Volume of BSA	Final BCA Concentration (µg/µl)
A	0	300 µl stock	2.000
B	125	375 µl stock	1.500
C	325	325 µl stock	1.000
D	175	175 µl of B	0.750
E	325	325 µl of C	0.500
F	325	325 µl of E	0.250
G	325	325 µl of F	0.125
H	400	100 µl of G	0.025
I	400	0 µl	0.000 = blank

For assay: 5 µl of each standard + 20 µl of ddH₂O = 25 µl total

- 5
 - b. Working reagents

Mix 50ml of Reagent A with 1ml of Reagent B

*The Working reagent is stable for several days when stored in a closed container at room temperature
- 10
 2. Preparation of samples in 96-well plate.

5 µl of purified protein (from step 20 of Purification procedure)

20 µl of ddH₂O

Mix well
- 15
 3. Assay procedure
 - a. Add 200 µl of Working Reagent to each well containing 25 µl of standards and samples
 - b. Mix plate thoroughly on a plate shaker for 30 seconds
 - c. Cover plate with aluminum foil tape
 - 20
 - d. Incubate at 37°C for 30 minutes
 - e. Cool plate to room temperature
 - f. Measure the absorbance at 562 nm on a plate reader
 4. Use Excel for standard curve plotting and determine protein concentration of samples
 - 25
 5. Normalize protein concentration for assay
 - a. Run a protein gel of normalized protein to confirm concentration

WO 02/090300

PCT/US02/14358

- b. Stain with SYPRO Orange for 30 min- 1 hr (and/or Coomassie blue overnight)
- c. Visualize gel on Alpha Innotech Corporation Imager
- d. Perform densitometry using Kodak 1D 3.5 Network software

5 THIOREDOXIN REDUCTASE ASSAY

1. Assay is set up in 384 microtiter plates with 50 μ l final volume per assay/well : Upto 4x96 well plate into one 384 plate, specific pattern to be noted at time of transfer.
 2. Transfer 5 μ l of normalized protein samples to 384 microtiter plate wells. NADPH or NADH at 1.2mM (or other appropriate concentrations), and 2 μ M of Purified Thioredoxin substrate is used in assay.
 3. Prepare assay mix:

	1rxn	300rxn
ddH ₂ O	35.1 μ l	10.53ml
1M Tris pH 8.0	5.0 μ l	1.5ml
0.5M EDTA	1.0 μ l	300 μ l
20mM DTNB	0.5 μ l	150 μ l
25mM NADPH or NADH	2.4 μ l	720 μ l
100 μ M Purified Thioredoxin	1 μ l	300 μ l
Total	45 μ l	13.5 ml
 - *Add NADH or NADPH and Thioredoxin substrate immediately before adding assay mix to supernatant to be tested
 4. Use Titertek Multidrop 384 to add 45 μ l of assay mix
 5. Immediately place plate in Spectramax plate reader to begin data collection
 6. For measurement of kinetic parameters (Kcat and Km) the following substrate concentration ranges were generally used:

NADPH : 0.00, 0.01, 0.02, 0.04, 0.08, 0.15, 0.3, 0.6, 1.2, 2.5, 5.0 & 10.0 mM

NADH: 0.02, 0.04, 0.08, 0.15, 0.3, 0.6, 1.2, 2.5, 5.0, 10.0 & 20.0 mM.
- Initial reaction rate in the linear range was determined for each concentration. The data was analyzed using GraphPad Prism software to fit a standard Michaelis-Menton equation.

Preparation of Thioredoxin h (N terminal His Tag) for Assay Use

Culture preparation:

1. Inoculate 2 liter expression culture with overnight culture of Thioredoxin-codon opt.e-coli/pET28b in BL21 Star (DE3) expression cells. This yields > 100 mgs of purified protein.
2. After growth period, induce cells with 1M IPTG for a final concentration of 1mM IPTG. Grow overnight at 30°C, 250rpm.

WO 02/090300

PCT/US02/14358

3. Next day, spin down the 2L culture into 20 50ml Falcon tubes and discard the supernatant leaving just the pellet from 100ml of culture. Freeze pellets at -80°C before continuing with supernatant preparation and His-tag purification.

Supernatant preparation:

- 5
 1. Resuspend 20 pellets in 1ml Bugbuster each and shake at 250rpm, room temperature for 20min.
 2. Spin down cells and combine supernatants into a 50ml Falcon tube. Add equal volume of 2X Loading buffer with 2-mercaptoethanol. Proceed with purification.
- 10 His-tag protein purification:
 1. Add 6ml Clontech TALON Superflow resin suspension to four 50ml Falcon tubes.
 2. Wash resin with 30ml of 1X Loading buffer twice
 3. Bind protein to resin by gently agitating at room temperature for 20min.
 4. Wash resin in 30ml of 1X Loading buffer at room temperature for 10min.
- 15
 5. Resuspend resin in 3ml of 1X Loading buffer.
 6. Combine suspensions from all four tubes into one Clontech 10ml gravity flow column.
 7. Wash resin with 15ml of 1X Loading buffer.
 8. Resuspend resin in 20ml of 250mM imidazole elution buffer. Elute protein into a 50ml tube twice.
- 20
 9. Continue with imidazole removal by filtration and sample concentration or freeze at -20°C for later use.

Filtration and concentration of purified thioredoxin:

1. Run purified protein sample through Millipore Ultrafree-4 Biomax 5K filter tubes.
- 25
 2. Wash samples three times with Filtration Wash buffer.
 3. Combine concentrated protein samples together. Perform a BCA assay to determine concentration and then dilute to 100uM with 50% glycerol, 20mM Tris-HCl pH 8.0.

2X Loading buffer

- 30
 - 100mM NaPO₄ pH 8.0
 - 10mM Tris, pH 8.0
 - 600mM NaCl
 - 20mM Imidazole
 - 10% Ethylene glycol

- 35 For 2X Loading buffer with 2mM 2-mercaptoethanol, add 0.156ul/ml

250mM imidazole elution buffer50mM NaPO₄ pH 8.0

WO 02/090300

PCT/US02/14358

5mM Tris, pH 8.0
 200mM NaCl
 250mM Imidazole
 10% Ethylene glycol

5

Filtration Buffer (for imidazole removal)

50mM NaPO₄
 10mM Tris, pH 8.0
 200mM NaCl
 10% Ethylene glycol
 ddH₂O

10

15

Example 2

Transformation of Plants with Variant TR proteins

Overview

A gene encoding an oleosin-TR fusion protein, an oleosin-TR-reductase fusion protein or an oleosin-hybrid TR-reductase/TR-reductase fusion protein can be incorporated into plant cells using conventional recombinant DNA technology. Generally, this involves inserting a DNA molecule encoding an oleosin-TR-reductase fusion protein, an oleosin-TR-reductase fusion protein or an oleosin-hybrid TR/TR-reductase fusion protein into an expression system as described above.

20

Breeding

25

Plants expressing an oleosin-TR fusion protein, an oleosin-TR-reductase fusion protein or an oleosin-hybrid TR/TR-reductase fusion protein, in combination with other characteristics important for production and quality, can be incorporated into plant lines through breeding approaches and techniques known in the art. Where a plant expressing an oleosin-TR fusion protein, an oleosin-TR-reductase fusion protein or an oleosin-hybrid TR/TR-reductase fusion protein is obtained, the transgene is moved into commercial varieties using traditional breeding techniques without the need for genetically engineering the allele and transforming it into the plant.

30

Plants having the capacity for apomictic reproduction, in which maternal tissue gives rise to offspring, can be transformed to express an oleosin-R fusion protein, an oleosin-TR-reductase fusion protein or an oleosin-hybrid TR/TR-reductase fusion protein, and the introduced alleles can be maintained in desired backgrounds by apomictic breeding.

35

Isolation of TR and TR-reductase genes and in vitro assays

WO 02/090300

PCT/US02/14358

In one embodiment, TR genes from *Arabidopsis*, wheat, a mammalian source such as calf and *E. coli* can be isolated and expressed in *E. coli* using bacterial expression vectors, and the resulting protein product can be purified. In another embodiment, TR-reductase genes from *Arabidopsis* and *E. coli* can be isolated, expressed in *E. coli*, and purified. In addition, the TR/TR-reductase gene can be isolated/obtained from *Mycobacterium leprae* and expressed in *E. coli* and purified. In a preferred embodiment, *M. leprae* codons may be altered for optimization in any given host, such as an *E. coli* host cell or a plant species. Codon usage tables for many organisms are known and available, permitting codon optimization of coding sequences tailored for a particular host.

- 10 In another embodiment TR-reductases with altered cofactor specificity are prepared using targeted mutagenesis or random mutagenesis, and tested for specific mutations at the cofactor binding site (Shiraishi, *et al.* (1998) *Arch Biochem Biophys* 358 (1): 104-115; Galkin *et al.* (1997) *Protein Eng* 10(6): 687-690; Carugo *et al.* (1997) *Proteins* 28(1):10-28; Hurley *et al.* (1996) *Biochemistry* 35(18):5670-8; and/or by addition of organic solvent (Holmberg *et al.* (1999) *Protein Eng* 12 (10): 851-856). Determination of mutations could be assisted by computer programs such as the one developed by Mayo and Dahiyat (*Chem & Eng News* October 6, 1997, pages 9-10). Each of the foregoing references is incorporated herein by reference in its entirety.

- 20 Combinations of different TRs and TR-reductases are used in a matrix to determine which TR and TR-reductase combination is most effective in the reduction of wheat storage proteins and milk storage protein β -lactoglobulin in vitro. Preferably, a combination of TR and TR-reductase are tested. These experiments are carried out as described in Del Val *et al.* ((1998) *Jnl Allerg Clin Immunol* 103:690-697). Inbred high-IgE-responder atopic dogs are obtained and further prepared by sensitization with commercial extracts of food preparations including milk and wheat. Skin tests are performed using the Type I hypersensitivity reaction. Evans blue dye is injected intravenously shortly before skin testing. Aliquots of wheat gruel, whole cow's milk extract and pure β -lactoglobulin are injected intradermally. Skin tests are read blindly by scoring 2 perpendicular diameters of each blue spot. The ability of oleosin-TR, oleosin-TR-reductase and combinations thereof to affect the allergic response is measured in the presence and absence of exogenous NADPH or NADH.

- 30 Construction of plant expression vectors

The *Arabidopsis* TR and TR-reductase gene sequences have been published (Rivera-Madrid *et al.* (1995) *Proc Natl Acad Sci USA* 92:5620-5624; Jacquot *et al.* (1994) *J Mol Biol* 235:1357-1363), and these genes can be isolated by PCR.

- 35 In one embodiment, both the *Arabidopsis* TR and TR-reductase genes are translationally fused to both the N- and C-terminal end of oleosin. This open reading frame is under transcriptional control of appropriate promoter and terminator sequences for expression in plants. In a preferred embodiment, the phaseolin promoter and terminator sequences are used to create *Arabidopsis* TR (ATR) and *Arabidopsis* TR-reductase (ATRR) constructs.

WO 02/090300

PCT/US02/14358

Expression in Arabidopsis

In one embodiment, *Arabidopsis* is used as a model system for the initial testing of oleosin-ATR and oleosin-ATRR expression constructs. Seed of *Arabidopsis* contain oleosin-coated oil bodies very similar to crop species, especially oilseed crop species, that can be used for commercial production of TR. Expression of oleosin-TR and oleosin-TR-reductase in *Arabidopsis* is used to obtain oleosin-TR and oleosin-TR-reductase fusions in oil bodies and to determine whether these fusion proteins are biologically active. Both N- and C-terminal fusions of both TR and TR-reductase to oleosin are made and tested. In a further embodiment, an oleosin fusion to the natural TR/TR-reductase fusion gene from *M. leprae* is tested. Accumulation of these fusion proteins is quantified using Western blotting, utilizing antibodies specific for oleosin and/or TR and TR-reductase. *Arabidopsis* is useful for this purpose since the time required to regenerate and grow transformed *Arabidopsis* plants and determine transgene expression and accumulation of expressed products in seeds is much shorter than for most crop species.

Construction of plant expression vectors

Plant expression vectors are constructed using other genes encoding TR and TR-reductase including, but not limited to, TR genes from wheat, TR genes from a mammalian source such as calf, the TR gene from *E. coli*; the TR-reductase gene from *E. coli*; and the TR/TR-reductase gene from *M. leprae*. Either or both of these genes are translationally fused to both the N and C-terminal end of oleosin. The open reading frame of any such construct is under the transcriptional control of appropriate promoter and terminator sequences. In a preferred embodiment, the phaseolin promoter and terminator sequences are used to construct plant expression vectors which are designated as TR' and TR-reductase. Even more preferably, the phaseolin promoter and terminator sequences are used to construct plant expression vectors which are designated as TR' and TR-reductase'.

Expression in Safflower

Plant transformation vectors as described above are used to transform safflower using methods known to those skilled in the art. In a preferred embodiment, safflower is transformed by a method adapted from the method disclosed by Baker and Dyer (*Plant Cell Rep* (1996) 16:106-110). Expression is assayed using Northern and Western blotting. The ability of the TR' and TR-reductase' constructs to reduce wheat storage proteins and milk storage protein β -lactoglobulin is tested. A minimum of 25 independently transformed transgenic safflower plants for each construct is generated. All the transgenic target crop plants are tested for oleosin-TR' and oleosin-TR-reductase' expression. The results from this analysis indicate which transformation event results in the highest and/or most optimal TR' or TR-reductase' activity. Transgenic lines transformed with this construct are subjected to further analyses. The quantity of TR' and TR-reductase' is determined using quantitative Western blotting analysis. The specific activity of the oleosin fusions is compared to the specific activity of the "free" TR' and TR-reductase' produced in *E. coli*.

WO 02/090300

PCT/US02/14358

Plant lines with the highest expression are propagated. Homozygotes and double haploid plants can be produced that possess a stable genotype to ensure stable transgene inheritance in subsequent generations.

5 Preparation of biotinylated TR

In one embodiment, TR can be biotinylated *in vitro* by chemical modification of the lysine residues using chemical agents such as biotiny-N-hydroxysuccinimide ester. As an alternate embodiment, an *in vivo*, site-specific biotinylation utilizing a biotin-domain peptide from the biotin carboxy carrier protein of *E. coli* acetyl-CoA carboxylase may be used as described by Smith *et al.* ((1998) *Nuc Acid*

- 10 *Res* 26:1414-1420). A recombinant thioredoxin capable of being biotinylated *in vivo* by the *E. coli* host endogenous biotinylation machinery (BIOTRX) is constructed by inserting an oligonucleotide encoding a 23 amino acid biotinylation recognition peptide in-frame at the 5'-end of *E. coli* *trxA*, creating the construct pBIOTRX. Cells containing the pBIOTRX plasmid are grown in the absence of exogenous biotin and the amount and solubility of BIOTRX protein is determined. Up to 10% of total cellular protein is found to be BIOTRX protein, while a low amount of tritiated biotin is incorporated
- 15 into BIOTRX protein and BIOTRX binding to immobilized avidin or immobilized avidin-alkaline-phosphatase is low. Addition of 10 µg/ml biotin to the pre-induction medium of pBIOTRX-transformed cells results in an improvement in the overall extent of biotin incorporation.

20 Preparation of biotinylated oil bodies-TR mixtures

Avidin or streptavidin are used to link the biotinylated TR to biotinylated oil bodies. Purified biotinylated TR is mixed with biotinylated oil bodies at different ratios. The efficacy of these mixtures to reduce allergenicity and improve dough quality in wheat is tested as well as the efficacy of these mixtures to reduce allergenicity in milk preparations. The controls include wild type safflower oil

- 25 bodies and wild type safflower oil bodies mixed, but not linked, with TR.

WO 02/090300

PCT/US02/14358

CLAIMS

We claim:

1. A method for altering the cofactor specificity of thioredoxin reductase comprising computational mutagenesis.
2. A method according to claim 1 for altering the cofactor specificity of thioredoxin reductase comprising:
 - a) inputting a set of coordinates for a thioredoxin reductase (TR) scaffold protein comprising amino acid positions,
 - b) applying at least one protein design cycle; and
 - c) generating a set of candidate variant proteins with altered cofactor dependency.
3. A method according to claim 2 wherein said TR scaffold proteins are selected from the group consisting of *E. coli*, *Bacillus subtilis*, *Mycobacterium leprae*, *Saccharomyces*, *Neurospora crassa*, *Arabidopsis*, and human.
4. A method according to claim 1 or 2 wherein said cofactor specificity of said variant TR is NADPH or NADH.
5. A method according to claim 1 or 2 wherein said cofactor specificity of said variant TR is switched to NADH.
6. A method according to claim 1 or 2 wherein said cofactor specificity of said variant TR is altered such that said variant preferentially binds NADPH compared to NADH.
7. A method according to claim 1 or 2 wherein said cofactor specificity of said variant TR is altered such that said variant preferentially binds NADH compared to NADPH.
8. A method according to claim 1 or 2 wherein said cofactor specificity of said variant TR is altered such that said variant exhibits improved catalytic efficiency for NADPH as compared to a wild-type TR protein.
9. A method for altering the substrate specificity of thioredoxin reductase comprising:
 - a) inputting a set of coordinates for a thioredoxin reductase scaffold protein comprising amino acid positions,
 - b) applying at least one protein design cycle; and
 - c) generating a set of candidate variant proteins with altered substrate specificity.

WO 02/090300

PCT/US02/14358

10. A variant thioredoxin reductase (TR) protein according to claim 9 wherein said variant TR protein reduces a thioredoxin protein obtained from an organism selected from the group consisting of *E. coli*, *Bacillus subtilis*, *Mycobacterium leprae*, *Saccharomyces*, *Neurospora crassa*, *Arabidopsis*, and human.
- 5 11. A variant TR protein according to claim 10 or 2, wherein said variant protein is fused to a second protein, wherein said second protein is either a wild-type TR protein, thioredoxin, or a variant TR protein.
- 10 12. The variant TR protein according to claim 11, wherein said variant protein is fused to said second protein through a linker.
13. A variant TR protein according to claim 1 or 2 wherein said variant TR protein has from 1 to 3 amino acid substitutions as compared to the wild-type *Arabidopsis* TR protein.
- 15 14. A variant TR protein according to claim 13 wherein said amino acid substitutions are selected from positions A4, A5 and A6.
15. A variant TR protein according to claim 14 wherein said amino acid substitutions are selected from the group of substitutions consisting of RA4W, RA5L, RA5M, RA5I, RA5F, RA5V, RA5Y, RA6T, RA6S, RA6Q, RA6G, and RA6N.
- 20 16. A variant TR protein according to claim 15 comprising the amino acid substitutions RA4W and RA6T.
- 25 17. A variant TR protein according to claim 15 comprising the amino acid substitutions RA4W, RA5L, and RA6S.
18. A variant TR protein according to claim 15 comprising the amino acid substitutions RA5Y and RA6N.
- 30 19. A variant TR protein according to claim 15 comprising the amino acid substitutions RA4W, RA5F, and RA6Q.
- 35 20. A method for altering the cofactor specificity of target protein comprising:
- a) inputting a set of coordinates for a scaffold protein comprising amino acid positions,
 - b) applying at least one protein design cycle; and
 - c) generating a set of candidate variant proteins with altered cofactor specificity.

WO 02/090300

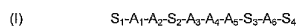
PCT/US02/14358

21. A method according to claim 1, 2, 9 or 20 wherein said protein design cycle comprises protein design automation (PDA™).

22. A method according to claim 1, 2, 9 or 20 wherein said protein design cycle comprises the sequence prediction algorithm.

23. A method according to claim 1, 2, 9 or 20 wherein said protein design cycle comprises a force field calculation.

24. A variant thioredoxin reductase (TR) protein comprising an isolated polypeptide molecule of Formula I



wherein

- a) S_1 comprises a polypeptide sequence selected from the group consisting of SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, and SEQ ID NO:7, or a sequence having substantial similarity thereto;
- b) S_2 comprises a polypeptide sequence selected from the group consisting of SEQ ID NO:8, SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, and SEQ ID NO:14, or a sequence having substantial similarity thereto;
- c) S_3 comprises a polypeptide sequence selected from the group consisting of SEQ ID NO:15, SEQ ID NO:16, SEQ ID NO:17, SEQ ID NO:18, SEQ ID NO:19, SEQ ID NO:20, and SEQ ID NO:21, or a sequence having substantial similarity thereto;
- d) S_4 comprises a polypeptide sequence selected from the group consisting of SEQ ID NO:22, SEQ ID NO:23, SEQ ID NO:24, SEQ ID NO:25, SEQ ID NO:26, SEQ ID NO:27, and SEQ ID NO:28, or a sequence having substantial similarity thereto;
- e) A_1 is an amino acid moiety selected from the group consisting of serine, valine, glycine, alanine, leucine, isoleucine, methionine, phenylalanine, and tryptophan;
- f) A_2 is an amino acid moiety selected from the group consisting of alanine, glycine, valine, leucine, isoleucine, methionine, phenylalanine, and tryptophan;
- g) A_3 is an amino acid moiety selected from the group consisting of histidine, aspartic acid, glutamic acid, arginine, leucine, serine, threonine, cysteine, asparagine, glutamine, and tyrosine;
- h) A_4 is an amino acid moiety selected from the group consisting of arginine, alanine, glycine, valine, leucine, isoleucine, methionine, phenylalanine, and tryptophan;
- i) A_5 is an amino acid moiety selected from the group consisting of arginine, asparagine, glutamine, aspartic acid, glutamic acid, cysteine, serine, threonine, and lysine;
- j) A_6 is an amino acid moiety selected from the group consisting of arginine, glutamic acid, asparagine, glutamine, aspartic acid, cysteine, serine, threonine, and lysine;

WO 02/090300

PCT/US02/14358

provided that at least

- 5 A₁ is not serine;
 A₂ is not alanine;
 A₃ is not histidine;
 A₄ is not arginine;
 A₅ is not arginine; or
 A₆ is not arginine.

- 10 25. The polypeptide molecule according to claim 24, wherein S₁ consists of a polypeptide sequence having the sequence selected from the group consisting of SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, and SEQ ID NO:7.
- 15 26. The polypeptide molecule according to claim 24, wherein S₂ consists of a polypeptide sequence having the sequence selected from the group consisting of SEQ ID NO:8, SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, and SEQ ID NO:14.
- 20 27. The polypeptide molecule according to claim 24, wherein S₃ consists of a polypeptide sequence having the sequence selected from the group consisting of SEQ ID NO:15, SEQ ID NO:16, SEQ ID NO:17, SEQ ID NO:18, SEQ ID NO:19, SEQ ID NO:20, and SEQ ID NO:21.
- 25 28. The polypeptide molecule according to claim 24, wherein S₄ consists of a polypeptide sequence having the sequence selected from the group consisting of SEQ ID NO:22, SEQ ID NO:23, SEQ ID NO:24, SEQ ID NO:25, SEQ ID NO:26, SEQ ID NO:27, and SEQ ID NO:28.
- 30 29. The polypeptide molecule according to claim 24, wherein S₁ is the polypeptide sequence set forth in SEQ ID NO:1, S₂ is the polypeptide sequence set forth in SEQ ID NO:8, S₃ is the polypeptide sequence set forth in SEQ ID NO:15, and S₄ is the polypeptide sequence set forth in SEQ ID NO:22.
- 35 30. The polypeptide molecule according to claim 24, wherein S₁ is the polypeptide sequence set forth in SEQ ID NO:2, S₂ is the polypeptide sequence set forth in SEQ ID NO:9, S₃ is the polypeptide sequence set forth in SEQ ID NO:16, and S₄ is the polypeptide sequence set forth in SEQ ID NO:23.
31. The polypeptide molecule according to claim 24, wherein S₁ is the polypeptide sequence set forth in SEQ ID NO:3, S₂ is the polypeptide sequence set forth in SEQ ID NO:10, S₃ is the polypeptide sequence set forth in SEQ ID NO:17, and S₄ is the polypeptide sequence set forth in SEQ ID NO:24.
32. The polypeptide molecule according to claim 24, wherein S₁ is the polypeptide sequence set forth in SEQ ID NO:4, S₂ is the polypeptide sequence set forth in SEQ ID NO:11, S₃ is the polypeptide sequence set forth in SEQ ID NO:18, and S₄ is the polypeptide sequence set forth in SEQ ID NO:25.

WO 02/090300

PCT/US02/14358

33. The polypeptide molecule according to claim 24, wherein S_1 is the polypeptide sequence set forth in SEQ ID NO:5, S_2 is the polypeptide sequence set forth in SEQ ID NO:12, S_3 is the polypeptide sequence set forth in SEQ ID NO:19, and S_4 is the polypeptide sequence set forth in SEQ ID NO:26.
- 5 34. The polypeptide molecule according to claim 24, wherein S_1 is the polypeptide sequence set forth in SEQ ID NO:6, S_2 is the polypeptide sequence set forth in SEQ ID NO:13, S_3 is the polypeptide sequence set forth in SEQ ID NO:20, and S_4 is the polypeptide sequence set forth in SEQ ID NO:27.
- 10 35. The polypeptide molecule according to claim 24, wherein S_1 is the polypeptide sequence set forth in SEQ ID NO:7, S_2 is the polypeptide sequence set forth in SEQ ID NO:14, S_3 is the polypeptide sequence set forth in SEQ ID NO:21, and S_4 is the polypeptide sequence set forth in SEQ ID NO:28.
- 15 36. The polypeptide molecule according to claim 24, wherein A_1 is an amino acid moiety selected from the group consisting of valine, alanine, and leucine.
37. The polypeptide molecule according to claim 24, wherein A_2 is an amino acid moiety selected from the group consisting of glycine, valine, and leucine.
- 20 38. The polypeptide molecule according to claim 24, wherein A_3 is an amino acid moiety selected from the group consisting of aspartic acid, glutamic acid, asparagine, and glutamine.
39. The polypeptide molecule according to claim 24, wherein A_4 is an amino acid moiety selected from the group consisting of alanine, glycine, valine, leucine, isoleucine, and methionine.
- 25 40. The polypeptide molecule according to claim 24, wherein A_5 is an amino acid moiety selected from the group consisting of asparagine, glutamine, aspartic acid, and glutamic acid.
41. The polypeptide molecule according to claim 24, wherein A_6 is an amino acid moiety selected from the group consisting of glutamic acid, glutamine, aspartic acid, and asparagine.
- 30 42. The polypeptide molecule according to claim 24, wherein A_1 is valine.
43. The polypeptide molecule according to claim 24, wherein A_2 is glycine.
- 35 44. The polypeptide molecule according to claim 24, wherein A_3 is aspartic acid.
45. The polypeptide molecule according to claim 24, wherein A_4 is alanine.
46. The polypeptide molecule according to claim 24, wherein A_5 is asparagine.

WO 02/090300

PCT/US02/14358

47. The polypeptide molecule according to claim 24, wherein A₆ is glutamic acid.
48. The polypeptide molecule according to claim 24, wherein said molecule reduces a thioredoxin protein obtained from an organism selected from the group consisting of *E. coli*, *Bacillus subtilis*, *Mycobacterium leprae*, *Sarcocharomyces*, *Neurospora crassa*, *Arabidopsis*, and Human.
49. The polypeptide molecule according to claim 24, wherein said reduction of thioredoxin takes place in the presence of a co-factor.
50. The polypeptide molecule according to claim 24, wherein said co-factor is NADPH or NADH.
51. The polypeptide molecule according to claim 24, wherein said co-factor is NADH.
52. The polypeptide molecule according to claim 24, wherein said polypeptide shows greater than 100 times more affinity for NADPH than for NADH.
53. The polypeptide molecule according to claim 24, wherein said polypeptide shows greater than 50 times more affinity for NADPH than for NADH.
54. The polypeptide molecule according to claim 24, wherein said polypeptide shows greater than 25 times more affinity for NADPH than for NADH.
55. The isolated polypeptide molecule according to claim 24, wherein said polypeptide is fused to a second polypeptide.
56. The polypeptide molecule according to claim 55, wherein said polypeptide is fused to said second polypeptide through a linker.
57. The polypeptide molecule according to claim 56, wherein said linker comprises a polypeptide sequence having between about 5 and about 50 amino acids.
58. The polypeptide molecule according to claim 56, wherein said linker comprises a polypeptide sequence having between about 10 and about 40 amino acids.
59. The polypeptide molecule according to claim 56, wherein said linker comprises a polypeptide sequence having between about 15 and about 25 amino acids.
60. The polypeptide molecule according to claim 56, wherein said second polypeptide is thioredoxin.

WO 02/090300

PCT/US02/14358

61. The polypeptide molecule according to claim 56, wherein said polypeptide is further fused to a third polypeptide.
- 5 62. The polypeptide molecule according to claim 56 wherein said polypeptide is fused to said third polypeptide through a linker.
63. The polypeptide molecule according to claim 56 or 62, wherein said linker comprises a polypeptide sequence having a molecular weight between about 5 and about 100 kDa.
- 10 64. The polypeptide molecule according to claim 56 or 62, wherein said linker comprises a polypeptide sequence having a molecular weight between about 20 and about 70 kDa.
65. The polypeptide molecule according to claim 56 or 62, wherein said linker comprises a polypeptide sequence having a molecular weight between about 25 and about 45 kDa.
- 15 66. The polypeptide molecule according to claim 56 or 62, wherein said third polypeptide is oleosin.
67. A method for producing a plant with an a modified TR protein comprising:
- 20 (a) introducing an expression cassette comprising a promoter functional in a plant cell operably linked to a DNA molecule encoding a modified thioredoxin reductase (TR) protein according to claim 1 or 22 comprising an amino terminal chloroplast transit peptide, into the cells of a plant so as to yield transformed plant cells; and
- 25 (b) regenerating said transformed plant cells to provide a differentiated transformed plant, wherein expression of the DNA molecule encoding the modified TR protein in said plant alters the co-factor specificity compared to the untransformed plant.
- 30 68. A method according to claim 67 wherein said transformed plant expresses a modified TR protein wherein said cofactor specificity is NADPH or NADH.
69. A method according to claim 67 wherein said transformed plant expresses a modified TR protein wherein said cofactor specificity is switched to NADH.
- 35 70. A method according to claim 67 wherein said transformed plant expresses a modified TR protein wherein said cofactor specificity is altered such that said modified TR protein preferentially binds NADPH compared to NADH.

WO 02/090300

PCT/US02/14358

71. A method according to claim 1 or 2 wherein said transformed plant expresses a modified TR protein wherein said cofactor specificity is altered such that said modified TR protein exhibits improved catalytic efficiency for NADPH as compared to a wild-type TR protein in an untransformed plant.

5

72. A transformed plant prepared by the method of claim 67.

73. A transformed seed of said transformed plant of claim 72.

10 74. A method for making oil bodies comprising a modified thioredoxin reductase (TR) protein comprising:

- a) producing in a cell a modified TR protein according to claim 1 or 2;
- b) associating said modified TR protein with oil bodies through an oil body targeting protein capable of associating with modified TR protein and said oil bodies; and,
- 15 c) obtaining said oil bodies associated with said modified TR protein.

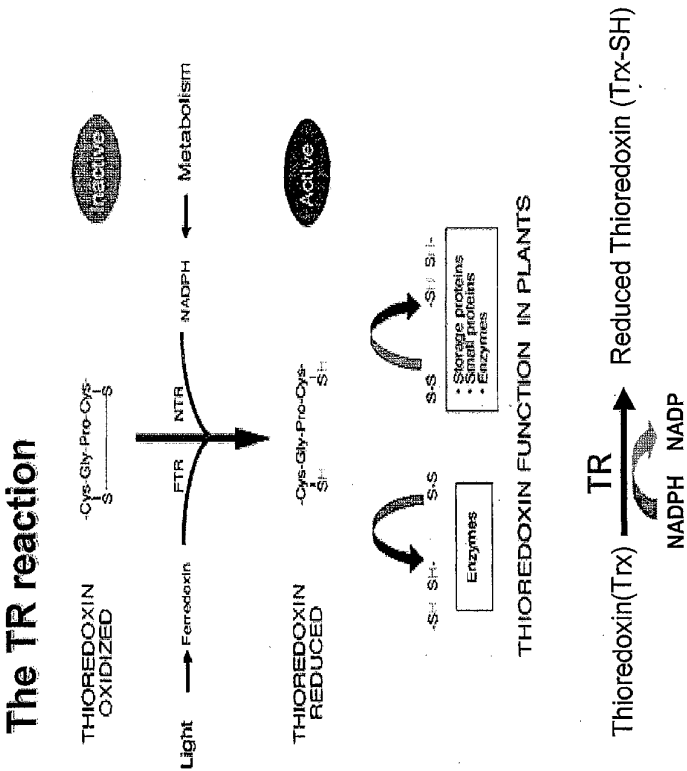
75. A method according to claim 74 further comprising:

- a) washing the oil bodies to obtain a washed oil body preparation comprising said modified TR protein; and,
- 20 b) formulating said washed oil bodies into an emulsion.

76. A method according to claim 74 wherein said oil bodies are used in the preparation of non-allergenic foods.

25 77. A method according to claim 74 wherein said oil bodies are used in the preparation of animal feeds to improve the digestibility of said feeds.

FIGURE 1



WO 02/090300

PCT/US02/14358

FIG 2A

THIOREDOXIN REDUCTASES

	A1	A2	S2	A3	A4	A5	S3	A6	species
Q39243	S170	A171	17X	H189	R190	R191	3X	R195	Arabidopsis
Q39242	S220	A221	17X	H239	R240	R241	3X	R245	Arabidopsis
O22229	T240	A241	17X	V259	R260	R261	3X	R265	Arabidopsis
P09625	T156	A157	17X	H175	R176	R177	3X	R181	E.coli
P29509	S164	A165	17X	V183	R184	R185	3X	R189	yeast
F38816	S188	A189	17X	V207	R208	R209	3X	R213	yeast
Q17745	V220	S221	17X	V239	R240	S241	3X	R245	C.elegans
Q9M2K1	V220	S221	17X	V239	R240	S241	3X	R245	C.elegans
Q9NMH3	V362	S363	17X	V381	R382	S383	3X	R387	C.elegans
Q9VNT5	V223	G224	17X	V242	R243	S244	3X	R248	Drosophila
O62768	V201	A202	17X	V220	R221	S222	3X	R226	bovine
Q9W278	V216	A217	17X	I235	R236	S237	3X	R241	bovine
Q16881	V201	A202	17X	V220	R221	S222	3X	R226	human
O95840	V229	A230	17X	M248	R249	S250	3X	R254	human
Q9UBS8	V201	A202	17X	V220	R221	S222	3X	R226	human
Q9UM79	V201	A202	17X	V220	R221	S222	3X	R226	human
Q9UQU8	V226	A227	17X	M245	R246	S247	3X	R251	human
Q9NMW6	V281	A282	17X	V300	R301	S302	3X	R306	human
Q9NNW7	V229	A230	17X	M248	R249	S250	3X	R254	human
Q9P101	V279	A280	17X	V298	R299	S300	3X	R304	human
Q9P210	V199	A200	17X	M218	R219	S220	3X	R224	human
Q9M225	V228	A229	17X	M247	R248	S249	3X	R253	human (mito)
Q99475	V253	A254	17X	V272	R273	S274	3X	R278	human
Q99P49	V315	A316	17X	V334	R335	S336	3X	R340	mouse
Q9CSV5	V201	A202	17X	V220	R221	S222	3X	R226	mouse
Q9CZ85	V317	G318	17X	V336	R337	S338	3X	R342	mouse
Q9JHA7	V229	A230	17X	M248	R249	S250	3X	R254	mouse
Q9JL74	V233	A234	17X	M252	R253	S254	3X	R258	mouse
Q9JMH5	V225	A226	17X	M244	R245	S246	3X	R250	mouse
Q9JMH6	V201	A202	17X	V220	R221	S222	3X	R226	mouse
O89049	V201	A202	17X	V220	R221	S222	3X	R226	rat
Q9JKE4	V201	A202	17X	V220	R221	S222	3X	R226	rat
Q9RL13	V201	A202	17X	V220	R221	S222	3X	R226	rat
Q9Z0J5	V231	A232	17X	M250	R251	S252	3X	R256	rat
Q9MYT8	V201	A202	17X	V220	R221	S222	3X	R226	pig

FIG 2B

GLUTATHIONE REDUCTASES

A1	A2	S2	A3	A4	A5	S3	A6	species	
P42770	I271	A272	17X	I290	R291	Q292	4X	R297	Arabidopsis
P48641	I214	A215	17X	F233	R234	K235	4X	R240	Arabidopsis
P48642	I211	A212	17X	Y230	R231	K232	4X	R237	rice
O64409	I83	A84	17X	I102	R103	Q104	4X	R109	maize
P06715	I178	A179	17X	V197	R198	K199	4X	R204	E.coli
O01412	I182	A183	17X	I201	R202	K203	4X	W208	O.volvulus
P41921	I208	G209	17X	I227	R228	G229	4X	R234	yeast
P91938	I192	G193	16X	V210	R211	S212	3X	R216	Drosophila
P00390	I242	A243	17X	I261	R262	H263	4X	R268	human
P47791	I220	A221	17X	I239	R240	H241	4X	R246	mouse
P70619	I136	A137	17X	I155	R156	H157	4X	R162	rat

FIGURE 2 A-E

WO 02/090300

PCT/US02/14358

FIG 2C

THIOREDOXIN REDUCTASES CONSENSUS:									
	A1	A2	S2	A3	A4	A5	S3	A6	
Arabidopsis	S(T)	A	17X	H(V)	R	R	3X	R	
S.coli	T	A	17X	H	R	R	3X	R	
yeast	S	A	17X	V	R	K	3X	R	
C.elegans	V	S	17X	V	R	S	3X	R	
Drosophila	V	G	17X	V	R	S	3X	R	
animals	V	A(S)	17X	V/H(I)	R	S	3X	R	

FIG 2D

GLUTATHIONE REDUCTASES CONSENSUS:									
	A1	A2	S2	A3	A4	A5	S3	A6	
plants	I	A	17X	I(P,V)	R	K/Q	4X	R	
bacteria	I	A	17X	V(I)	R	K	4X	R(W)	
yeast	I	G	17X	I	R	G	4X	R	
Drosophila	I	G	16X	V	R	S	3X	R	
animals	I	A	17X	I	R	H	4X	R	

FIG 2E

COFACTOR SPECIFICITY:									
	A1	A2	S2	A3	A4	A5	S3	A6	species . specificity
TR	S	A	17X	H	R	R	3X	R	Arabidopsis NADPH
TR	T	A	17X	H	R	R	3X	R	E.coli NADPH
GR(wt)	I	A	3XA13X	V	R	K	H3X	R	E.coli NADPH
GR(mut)	I	G	3XG13X	E	M	F	D3X	F	E.coli NADH
Op34	S	A	17X	H	Q	F	3X	Q	C.pasteurianum NADH
AhpF	S	G	17X	E	F	A	3X	K	S.typhimurium NADH

WO 02/090300

PCT/US02/14358

FIG 3A

SEQ ID NO:1

The N-terminus sequence (S₁) of *E. coli* thioredoxin reductase:

GTTHKSKLLILGSGPAGYTAAVYAAARANLQPVLITGMEKGGQLTTTTEVENWPGDPNDLT
GPLLMERMHEHATKFETELIFDHINKVDLQNRPFRLNGDNGETTCDALIIATGASARYLG
LPSEEAFFKGRGVSACATCDGFFYRNQKVAVIGGGN

FIG 3B

SEQ ID NO:2

The N-terminus sequence (S₁) of *Bacillus subtilis* thioredoxin reductase:

SHEKIYDVIIIGAGPAGMTAAVYTSRANLSTLMIERGIPGQGMANTEDVENYPGFESILG
PELSNMFEBHAKKFGARYAYGDIKEVIDGKEYKVVKAGSKEYKARAVIIAAGAEYKKIGV
PGEKELGGRGVSYCAVCDGAFFKGGELVVVGGGD

FIG 3C

SEQ ID NO:3

The N-terminus sequence (S₁) of *Mycobacterium leprae* thioredoxin reductase:

MNTTPSAHETIHEVIVIGSGPAGYTAALYAARQLTFLVFEGTSFGGALMTTTEVENYPG
FRNGITGPELMDDMREQALRFGAELRTEDVESVSLRGPIKSVVTABGQTYQARAVILAMG
TSVRYLQIFGEQELLGRGVSACATCDGSFRCQDIAVIGGGD

FIG 3D

SEQ ID NO:4

The N-terminus sequence (S₁) of *Saccharomyces* thioredoxin reductase:

VHNKVTIIIGSGPAAHTAAIYLARAEIKPILYEGMMANGIAAGGQLTTTTEIENFPGFDPG
LTGSELMDRMREQSTKFGPEIITETVSKVDLSKPKLWTFENEDABPVTTDAIILATGA
SAKRMHLPGEETYWQKGISACAVCDGAVPIFRNKPLAVIGGGD

FIG 3E

SEQ ID NO:5

The N-terminus sequence (S₁) of *Neurospora crassa* thioredoxin reductase:

MHSKVVIIGSGPAAHTAAIYLARAEIKPVLYEGFMANGIAAGGQLTTTTEIENFPGFDPG
IMGQELMDKMKASERFGTQIISETVAKVDLSARPFKYATEWSPEEYHTADSIILATGAS
ARRLHLPGEETYWQNGISACAVCDGAVPIFRNKHLVVIGGGD

FIGURES 3A–3BB

WO 02/090300

PCT/US02/14358

FIG 3F

SEQ ID NO:6

The N-terminus sequence (S₁) of *Arabidopsis* thioredoxin reductase:

MNGLETHNTRLCIVGSGPAAHTAAIYAARAEKPLLFEGWMANDIAPGGQITTTTDDVENF
PGFPEGILGVELTDKFRKQSERFGTTIFTETVTKVDFSSKPFKLFTDSKAILADAVILAT
GAVAKRLSPVGSGEASGGFWNRGISACAVCDGAAPIFRNKPLAVIGGGD

FIG 3G

SEQ ID NO:7

The N-terminus sequence (S₁) of Human thioredoxin reductase:

MNGPEDLFKSYDYDLIIIGGGSGGLAAKEAAQYGGKVMVLDFTPTPLGTRWGLGGTCV
NVGCIPKKLMHQALLGQALQDSRNYGKWVEETVKHDWDRMIEAVQNHIGSLNWGYRVAL
REKKVVYENAYGQFIGPHRIKATNNKGKEKIYSAESFLIATGERPRYLGPDKKEYCISS
DDLFSLLPYCPGKTLVVGASYVALECAGFLAGIGLGV

FIG 3H

SEQ ID NO:8

The first internal sequence (S₂) of *E. coli* thioredoxin reductase:

VEEALYLSNIASEVHLI

FIG 3I

SEQ ID NO:9

The first internal sequence (S₂) of *Bacillus subtilis* thioredoxin reductase:

VEEGVYLTRFASKVTIV

FIG 3J

SEQ ID NO:10

The first internal sequence (S₂) of *Mycobacterium leprae* thioredoxin reductase:

MEEALFLTRFARSVTLV

FIG 3K

SEQ ID NO:11

The first internal sequence (S₂) of *Saccharomyces* thioredoxin reductase:

CEEAQFLTKYGSKVFMFL

WO 02/090300

PCT/US02/14358

FIG 3L SEQ ID NO:12
The first internal sequence (S₂) of *Neurospora crassa* thioredoxin reductase:
AEEAMYLTKYGSHVTVL

FIG 3M SEQ ID NO:13
The first internal sequence (S₂) of *Arabidopsis* thioredoxin reductase:
MEEANFLTKYGSKVYII

FIG 3N SEQ ID NO:14
The first internal sequence (S₂) of Human thioredoxin reductase:
MVRSILLRGFDQDMANKIGEHMEEHGIFPI

FIG 3O SEQ ID NO:15
The second internal sequence (S₃) of *E. coli* thioredoxin reductase:
DGF

FIG 3P SEQ ID NO:16
The second internal sequence (S₃) of *Bacillus subtilis* thioredoxin reductase:
DKL

FIG 3Q SEQ ID NO:17
The second internal sequence (S₃) of *Mycobacterium leprae* thioredoxin reductase:
DEF

WO 02/090300

PCT/US02/14358

FIG 3R

SEQ ID NO:18
The second internal sequence (S₃) of *Saccharomyces* thioredoxin reductase:

DHL

FIG 3S

SEQ ID NO:19
The second internal sequence (S₃) of *Neurospora crassa* thioredoxin reductase:

DKL

FIG 3T

SEQ ID NO:20
The second internal sequence (S₃) of *Arabidopsis* thioredoxin reductase:

DAF

FIG 3U

SEQ ID NO:21
The second internal sequence (S₃) of Human thioredoxin reductase:

VPI

FIG 3V

SEQ ID NO:22
The C-terminus sequence (S₄) of *E. coli* thioredoxin reductase:

AEKILIKRLMDKVENGNIIILHTNRTLEEVTGDQMGVTGVRLRDTQNSDNIESLDVAGLFV
AIGHSPNTAIFEGQLELENGYIKVQSGIHGNATQTSIPGVFAAGDVMDHIYRQAITTSAGT
GCMALDAERYLDGLADAK

FIG 3W

SEQ ID NO:23
The C-terminus sequence (S₄) of *Bacillus subtilis* thioredoxin reductase:

AQSILQARAFDNEKVDPLWNKTVKEIHEENGKVGNTLVDTVTGEESEFKTDGVFIYIGM
LPLSKPFENLGITNEEGYIETNDRMETKVEGIFAAGDIREKSLRQIVTATGDGGSIAAQSV
QHYVBELQETLKTLLK

WO 02/090300

PCT/US02/14358

FIG 3X

SEQ ID NO:24
The C-terminus sequence (S₄) of *Mycobacterium leprae* thioredoxin reductase:

ASKIMLGRARNNDKIKFITNHTVVAVNGYTTVTG-LRLRNITTGEEITLVVTG

FIG 3Y

SEQ ID NO:25
The C-terminus sequence (S₄) of *Saccharomyces* thioredoxin reductase:

ASTIMQKRAEKNEKIEILYNTVALEAKGDGKLLNALRIKNTKKNEETDLPVSGLFYAIGH
TPATKIVAGQVDTDEAGYIKTVPGSSILTSVPGPFPAAGDVQDSKYRQAITSAGSGCMAALD
AEKYLTSLE

FIG 3Z

SEQ ID NO:26
The C-terminus sequence (S₄) of *Neurospora crassa* thioredoxin reductase:

ASSIMAHRLILNHEKVTVRFTNTVGVEVKGDDKGLMSHLVVKDVTGKEETLEANGLEYAIG
HDPATALVKQLETDADGYVVTKPGTILTSVEGVFAAGDVQDKRYRQAITSAGTGCMAAL
DAEKFLSEHEETPAEHRDTSVQGNLSTVKCDYENVFTTVFTPLEYGACGLSEKAVEKF
GEBNIEVYHSYFWPLEWTIPSRDNNKCYAKIICNTKDNBRVVGFHVLGPNAGEVTQGFPA
ALKCGLTKKQLDSTIGIHPVCAEVFTTLSVTKRSGASILQAGC

FIG 3AA

SEQ ID NO:27
The C-terminus sequence (S₄) of *Arabidopsis* thioredoxin reductase:

ASKIMQQRALSNPKIDVIWNSSVVEAYGGERDVLGGLKVKNVVTGDVSDLKVSGLFFAI
GHEPATKFLDGGVELDSDGYVVTKPGTTQTSVPGVFAAGDVQDKKYRQAITAAGTGCMAA
LDAEHYLQEIQSQQGKSD

FIG 3BB

SEQ ID NO:28
The C-terminus sequence (S₄) of Human thioredoxin reductase:

VRQIEAGTPGRLRVVAQSTNSEEIEGRYNTVMLAIGRDACTRKIGLETVGVKINEKTGK
IPVTDEEQTNVPYIYAI GDILEDKVELTPVAIQAGRLLAQRLYAGVVFVAGHEPRSSLVS
DVVDIDPDGYVLVKGRITSTSMGCVFAAGDLVDRTYRQAITAAGSGCAAAIDAERWLAEH
AGSKANETTEETGSDVDSITDTDWSTAMTDAKNAGVTIEVTDASFFADVLSNKPVLVDFW
ATWCGPCKMVAPVLEEIASEQRNQLTVAKLDVDTNPEMAREFCVVSIPTMILPQGGQPVK
RIVGAKKAALLRDLSDVVFNLN

WO 02/090300

PCT/US02/14358

An Overview of the HTP TR Screening

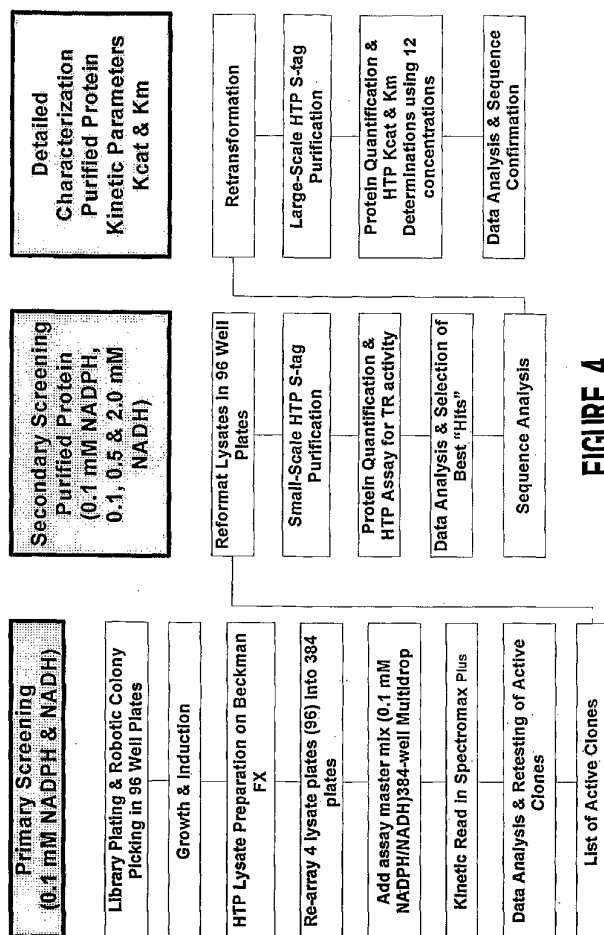


FIGURE 4

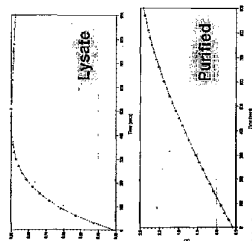
WO 02/090300

PCT/US02/14358

Protein Purification

HTP Strategies

Purified NTR/Trx
increases the dynamic
range of assay



More consistent
Kinetic data obtained

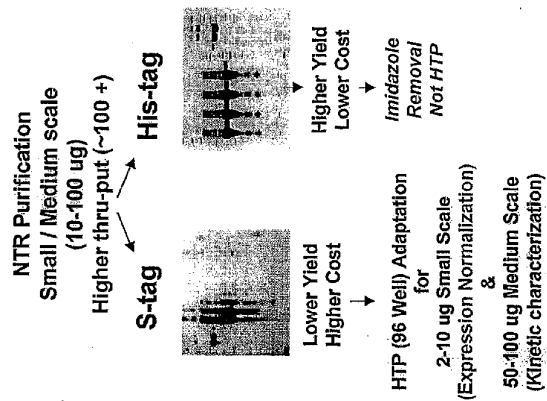
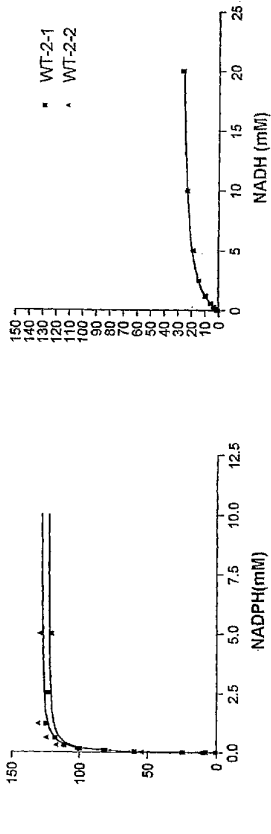


FIGURE 5

WO 02/090300

PCT/US02/14358

Kinetics of WT NTR with NAD(P)H

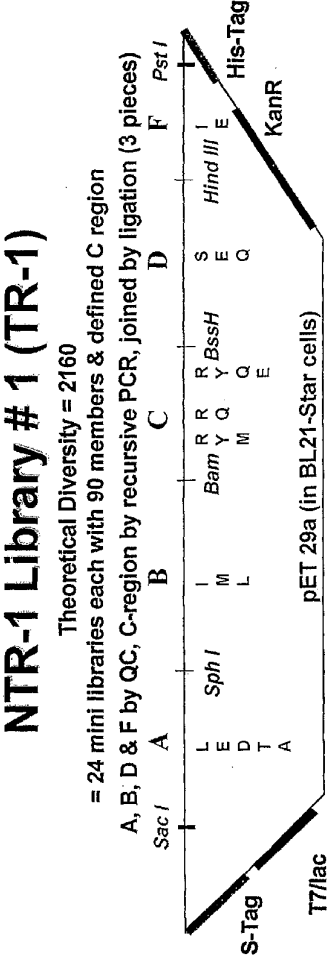


K_m (mM)	0.048	0.053	2.57	2.42
V_{max} (mOD/min)	123.8	128.9	29.5	29.0
	WT-2-1	WT-2-2	WT-2-1	WT-2-2

FIGURE 6

WO 02/090300

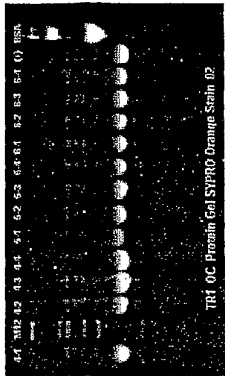
PCT/US02/14358



12/113

Sequence & Expression QC for TR-1

L-21 (25)	I-31 (38)	S-23 (30)	I-47 (60)
E-20 (24)	M-36 (44)	E-24 (30)	E-32 (40)
D-8 (10)	L-14 (18)	Q-32 (40)	
T-20 (24)			
A-15 (17)			
84	81	79	79



Expression...~90 % express based on 84 QC clones (7/84 NS)

Sequence diversity is well represented

FIGURE 7

WO 02/090300

PCT/US02/14358

NTR-1 Library # 2 (TR-2)

Theoretical Diversity = 2304
= 48 mini libraries each with 48 members & defined C region
AB & DF by QC, C-region by recursive PCR, joined by ligation (3 pieces)

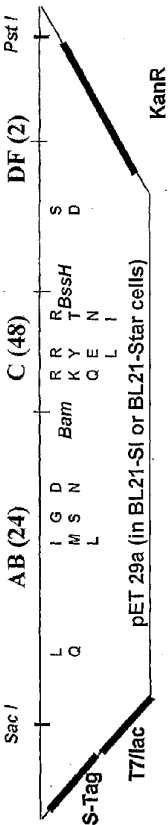


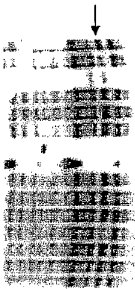
FIGURE 8

Sequence & Expression QC for TR-2

L 38	I 27	G 40	D 44	S 37
Q 46	M 33	S 45	N 41	D 49
		L 25		



Star vs SI expression



OW

>90 % of clones express soluble TR protein.

WO 02/090300

PCT/US02/14358

TR Library 1

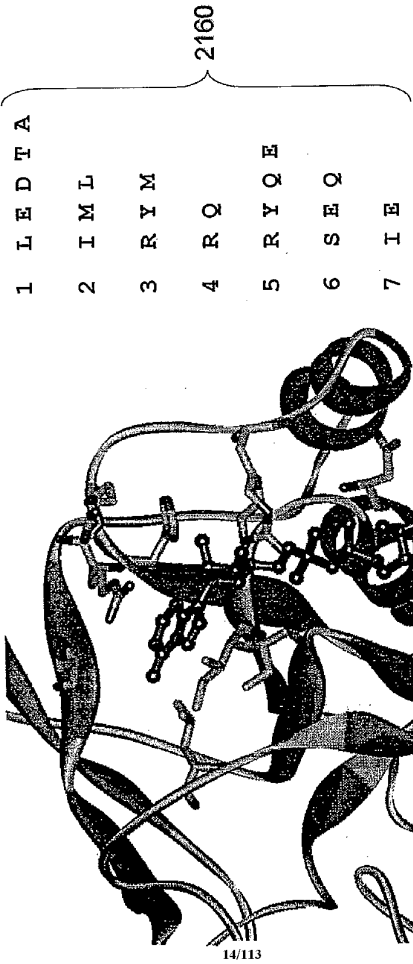
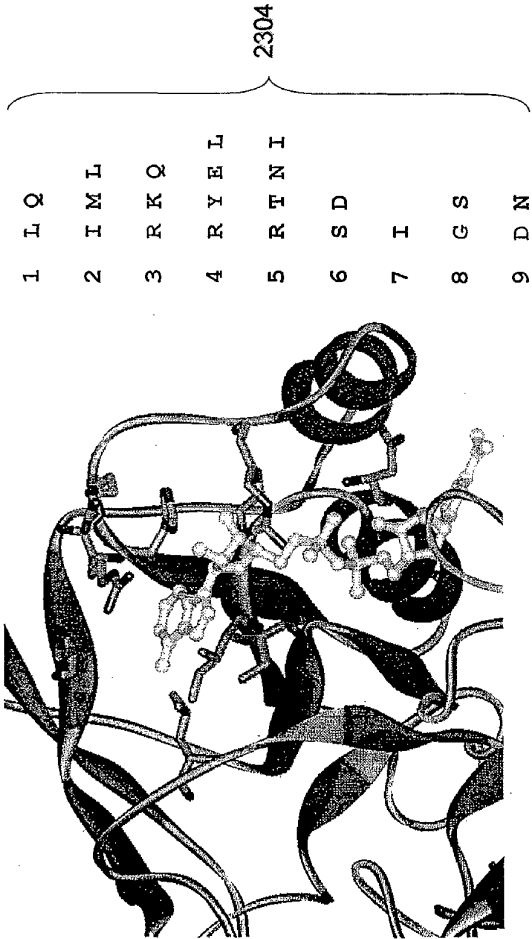


FIGURE 9A

WO 02/090300

PCT/US02/14358

TR Library 2



15/113

FIGURE 9B

The TR Libraries.....
FIGURE 10

WO 02/090300

PCT/US02/14358

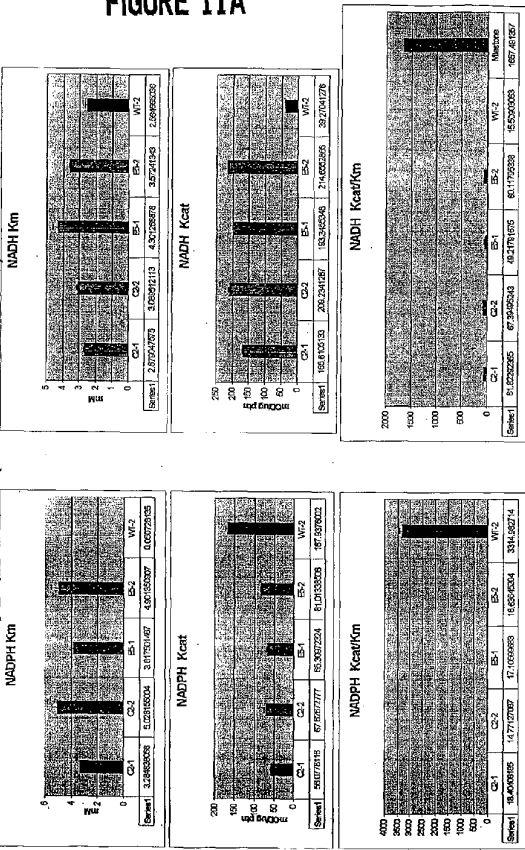
SacI AB BamHI C BssHII DF PstI				
Library	Design	1st Screen	2nd Screen	3rd Rd "Hits"
TR-1 (2160)	L I E M D L T A R R R Y Q M Y E S I E E Q	~4300	267	48 None
TR-2 (2304)	L I Q M L G D S N K Y T Q E N L I S D	~8600	569	34 3 +
TR-3 (18)	R R R M F N Y D	18	18	18 1 +
TR-4 (32)	A G H R D F Q C E L	16	16	16 None

WO 02/090300

PCT/US02/14358

FIGURE 11A

Kinetic Parameters of Two PDA™ designed & WT TR
Altered co-factor specificity & Improved catalytic efficiency.
Selectivity improvement (kcat of NADH/NADPH) of 1300%

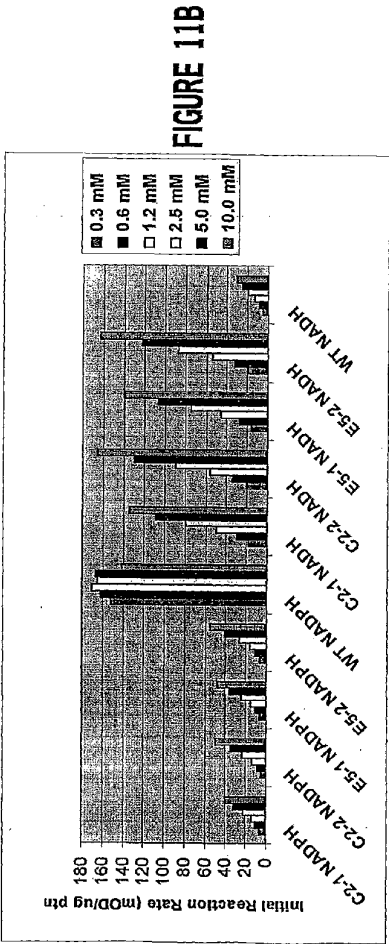


C2 is M-RYN and E5 is L-RYN, WT is I-RRR

WO 02/090300

PCT/US02/14358

Initial reaction rates of the thioredoxin reductase mutants at varying concentrations of co-factor



PDA™ designed NTRs have >50% of WT NADPH activity with NADH as co-factor.....!!!!



Purity & concentration check

Best Variants from TR-2 Design

Sp. Act. 1.2 mM	NADPH	NADH
RYN	25	106
RFN	24	108
RYN-A	19	97
RFN-A	15	81
WT-RRR	320	24
REN,RLN,RRN	65, 70, 340	65-75

FIGURE 12

WO 02/090300

PCT/US02/14358

FIGURE 13A

Activity Summary of Best "Hits" from New Library Designs

Sequenced Sample	0.6mM NADPH	1.2mM NADPH	0.6mM NADH	1.2mM NADH	1.2mM NADPH	1.2mM NADH	Protein (ug)	Sequenced Rank
RVN 38-2	62.79	94.89	30.71	50.84	117.35	63.67	1.52	RVN
WAN 22-2	5.74	29.59	29.59	44.89	12.32	55.99	1.49	WAN
WCT 13-2	5.25	9.23	31.15	54.84	10.11	65.85	1.75	WCT
WFQ 9-2	0.88	1.21	60.72	138.85	2.19	190.20	1.28	WFQ
WIS 38-2	53.98	87.75	83.01	144.47	111.75	191.05	1.19	WIS
WLG 28-1	11.91	19.28	50.86	89.90	25.74	103.99	1.33	WLG
WLS 28-2	19.35	50.90	74.51	127.79	41.62	155.55	1.41	WLS
WMD 4-2	0.57	0.65	94.83	55.28	1.08	77.70	1.27	WMD
WMG 21	18.32	31.20	72.03	116.75	37.87	138.09	1.36	WMG
WMS 6-1	18.63	92.71	65.90	119.21	38.18	131.51	1.46	WMS
WRG 25-2	322.04	329.94	46.84	79.09	425.71	97.90	1.50	WRG
WRM 28	110.82	181.98	35.42	61.25	228.34	74.89	1.46	WRM
WRS 19-2	327.73	310.02	61.72	88.49	375.23	123.23	1.28	WRS
WRT 31-1	255.30	250.52	93.35	146.43	370.72	185.01	1.39	WRT
WRT 31-2	303.21	328.56	105.15	155.22	397.01	197.60	1.26	WRT
WST 1-2	25.14	43.73	24.63	45.29	51.42	80.18	1.38	WST
WTS 16-2	40.95	71.02	23.51	44.73	83.42	47.16	1.34	WTS
WVG 15-2	59.72	107.69	16.83	92.51	134.90	110.87	1.12	WVG
WVR 20-2	475.28	593.27	72.03	128.98	731.63	157.60	0.86	WVR
WYS 9-2	31.70	93.72	36.23	67.11	72.94	75.93	1.19	WYS
RYN 50-1	8.18	14.27	34.91	57.12	16.92	68.01	1.30	RYN
RYN 51-1	7.74	13.25	27.32	51.05	17.46	61.66	1.30	RYN
RYN 52-1	9.97	16.53	33.52	57.54	21.43	77.32	1.46	RYN
RRR 54-2	206.78	277.81	10.09	17.18	287.75	19.00	1.38	RRR

The best hits are defined as having NADH Dependent activity better than RYN variant identified from TR-2

WO 02/090300

PCT/US02/14358

FIGURE 13B

Kinetic Parameters Summary of Best "Hits" from New Library Designs

Repeat

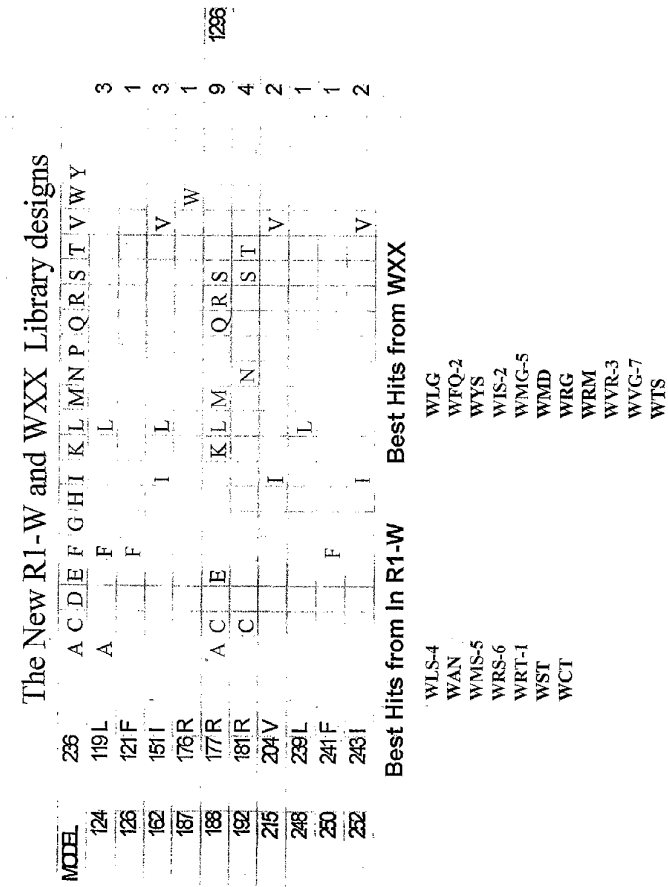
Sample	Vmax	Km
25-1	261.6	1.191
25-2	233.7	1.194
31-1	332.5	0.8293
31-2	363.8	0.821

Sample	Vmax	Km	Sequence 2	Protein In Assay (Vmax/Km)
38	207.1	3.749	RFQ	1.47 141.068443 37.6952676
42-1	69.53	4.274	RLE	0.63 103.6777271 25.69526234
42-2	250.1	4.227	RLE	1.671 155.1706003 36.70696206
39-1	268.3	2.594	RVN	1.561 191.1766042 73.70696564
39-2	314.7	2.682	RVN	1.52 207.591789 77.81975164
22-1	355.8	3.451	WAN	1.54 231.0797326 66.76975314
22-2	335.1	3.542	WAN	1.49 225.1612448 63.59595675
13-1	446.5	3.438	WGT	1.86 240.2100775 60.8691325
13-2	438.9	3.47	WGT	1.75 251.4353726 72.45990366
9-1	575.7	2.293	WFO	1.20 479.717094 209.2036904
9-2	650.1	2.283	WFO	1.29 509.8579269 223.4259115
33-1	565.5	2.291	WIS	1.26 464.0053622 202.5344226
33-2	530	2.352	WIS	1.13 460.9430657 199.7932092
28-1	522.5	3.335	WLG	1.33 384.216821 116.2180342
28-2	472.3	3.108	WLG	1.51 313.0195654 100.7141453
29-1	553	2.038	WLS	1.38 397.6218553 180.1871817
29-2	590.7	2.168	WLS	1.41 413.0369311 190.8069304
4-1	417.8	4.393	WMD	1.57 265.847776 69.34081381
4-2	412.2	4.331	WMD	1.27 324.0233635 66.5691627
21	527.8	2.272	WNG	1.36 385.73492 174.834966
6-1	577.6	2.439	WWS	1.46 395.3249718 162.4646834
25-1	592.4	1.953	WRG	1.43 211.9534595 132.4062411
25-2	528.7	1.951	WRG	1.50 219.3440385 132.3676763
26	373.9	2.674	WRN	1.46 235.554548 89.26741405
12-1	11.63	1.147	WRS	0.21 54.8235146 47.52267424
12-2	329.8	1.472	WRS	1.28 257.069876 174.7076041
31-1	407.8	0.896	WRT	1.38 295.5707186 326.865404
31-2	411.3	0.937	WRT	1.29 318.080573 340.0321228
1-2	295.5	4.093	WST	1.38 217.8100299 63.5081787
16-2	280.1	3.952	WTS	1.34 208.8005738 52.72780328
15-1	389	2.552	WVG	1.13 344.84770803 116.8181166
15-2	408.6	3.041	WVG	1.12 365.4524484 120.5083254
20-1	11.46	1.81	WVR	0.20 55.80985714 31.38610892
20-2	422.6	2.274	WVR	0.95 438.6732479 162.9062005
8-1	376.4	3.957	WYS	1.22 308.959202 77.52671727
8-2	362	3.693	WYS	1.19 304.2553633 82.33704666
50-1	220.8	1.794	RYN	1.30 189.4544076 94.4619192
50-2	240.9	2.473	RYN	1.30 194.8802844 74.7661652
53-1	297.3	2.095	RYN	1.45 205.4457183 88.077954
54-2	70.96	2.005	RYN	1.38 51.60534132 25.3382563

WO 02/090300

PCT/US02/14358

FIGURE 13C



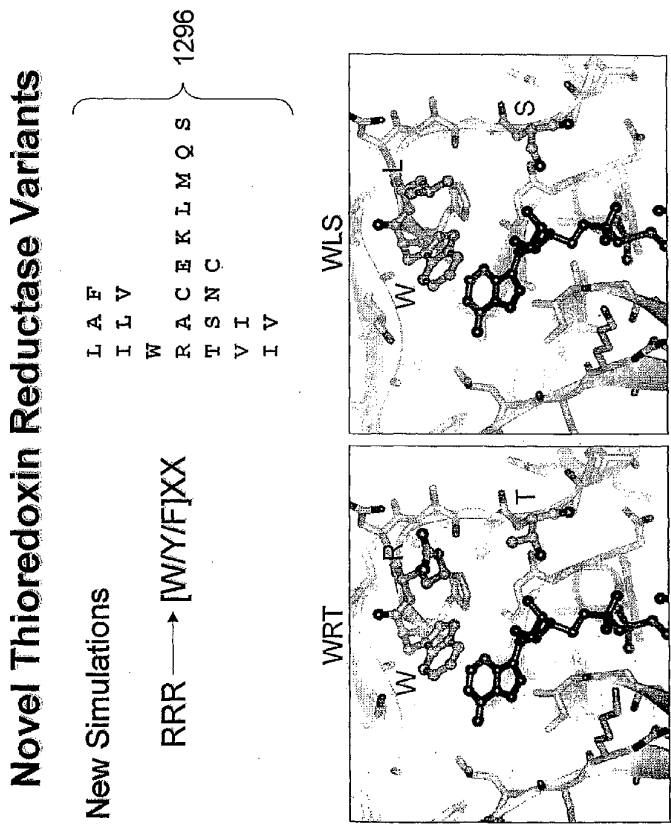


FIGURE 14

WO 02/090300

PCT/US02/14358

Diverse & Improved Catalytic Activities

- ✓Equal to or better than 50% of WT NADPH activity, with NADH at 1.2 mM.
- ✓Improved catalytic efficiency for the NADPH activity also.
- ✓13-fold better Kcat/Km and 2-fold lower Km for NADH compared to WT

Name	Experiment #1 (Average of duplicates)				Experiment #2 (Avg of quadruplicates)				Kinetic Parameters			
	0.5mM NADPH	1.2mM NADPH	3.0mM NADPH	1.2mM NADH	0.5mM NADPH	1.2mM NADPH	3.0mM NADPH	1.2mM NADH	Km	Kcat	Kcat/Km	P/H
RLW Library Hits	WRT	305.21	326.56	105.16	397.01	374.23	105.23	397.01	235.24	316.17	1.3263	
	WLS	19.35	30.90	74.61	41.62	27.33	65.64	41.62	2.17	419.04	190.89	
	WHS	18.63	32.71	65.50	38.16	131.61		38.16	2.44	398.32	162.49	
	WRS	327.73	316.02	61.75	375.23	123.29		375.23	1.67	257.17	174.71	
	WIS	53.99	67.75	83.01	111.75			111.75	2.35	155.52	159.76	
WXX Library Hits	WFQ	0.86	1.21	60.72	2.19	13.42		2.19	2.26	239.88	223.43	
	WVR	475.26	569.27	72.03	731.53	157.69		731.53	2.27	113.83	192.91	
	WVG	18.32	31.20	72.03	37.67	138.09		37.67	2.21	396.73	174.83	
	WVG	56.72	107.59	49.83	134.90	110.67		134.90	3.04	398.45	120.50	
	RYN	8.16	14.27	34.91	16.92	66.01		16.92	1.79	169.46	94.46	
TR-2	RRRWIT	209.76	127.83	10.09	19.00			19.00	2.01	64.61	25.74	
												High NADH-Dependent Activity
												High NADPH-Dependent Activity

FIGURE 15

24/113

WO 02/090300

PCT/US02/14358

The WVR variant coding sequence with S-tag at the N-terminus, His-Tag at C-terminus (5238-26)

```

1  tcgagcacca ccaccaccac cactgagatc cggctgctaa caaagcccca a
aggaagctg
61  agttggctgc tgccacgcgt gagcaataac tagcataacc ccttggggcc t
ctaaacggg
121 tottgagggg ttttttctg aaaggaggaa ctatatccgg attggcgaat g
ggacgcgcc
181 ctgtagcggc gcattaagcg cggcgggtgt ggtggttacg cgcagcgtga c
cgctacact
241 tgccagcgcc ctacgcgccg ctcttttcgc tttcttcctt tcctttctcg c
cacgttcgc
301 cggctttccc cgtcaagctc taaatcgggg gctcccttta gggttccgat t
tagtgcttt
361 acggcacctc gaccccaaaa aacttgatta gggtagtggt tcacgtagtg g
gccatcgcc
421 ctgatatagc gtttttcgcc ctttgacgtt ggagtccacg ttctttaata g
tggactctt
481 gttccaaaact ggaacaacac tcaaccctat ctcggtctat tcttttgatt t
ataagggat
541 tttgcgcgatt tcggcctatt ggttaaaaaa tgagctgatt taacaaaaat t
taacgcgaa
601 ttttaacaaa atattaacgt ttacaatttc aggtggcact ttccggggaa a
tgtcgcggg
661 aacccctatt tgtttatttt tctaaataca ttcaaatatg tatccgctca t
gaattaatt
721 cttagaaaaa ctcacgcagc atcaaatgaa actgcaattt attcatatca g
gattatcaa
781 taccatattt ttgaaaaagc cgtttctgta atgaaggaga aaactcaccg a
ggcagttcc
841 ataggatggc aagatcctgg tatcggctcg cgattccgac tcgtccaaca t
caatacaac
901 ctattaattt cccctcgtca aaaataaggt tatcaagtga gaaatcacca t
gagtgcaga
961 ctgaatccgg tgagaatggc aaaagtttat gcatttcott ccagacttgt t
caacaggcc
1021 agccattacg ctgctcatca aaatcactcg catcaaccaa accgttattc a
ttcgtgatt
1081 gcgcctgagc gagacgaaat acgcgatcgc tgttaaaagg acaattacaa a
cagggaatcg
1141 aatgcaaccg gcgcaggaaac actgccagcg catcaacaat attttcacct g
aatcaggat
1201 attcttctaa tacctggaat gctgttttcc cggggatcgc agtggtaggt a
accatgcat

```

FIGURE 16A

WO 02/090300

PCT/US02/14358

```

1261 catcaggagt acggataaaa tgcttgatgg tcggaagagg cataaattcc g
tcagccagt
1321 ttagtctgac catctcatct gtaacatcat tggcaacgct acctttgcca t
gtttcagaa
1381 acaactctgg cgcacgggc ttcccatata atcgatagat tgcgcgcact g
attgccga
1441 cattatcgcg agccattta taccatata aatcagcacc catgttgaa t
ttaatcgcg
1501 gcctagagca agcgtttcc cgttgaatat ggcataac accccttgta t
tactgttta
1561 tgtaagcaga cagttttatt gtcatgacc aaaatocctt aacgtgagtt t
tcgtccac
1621 tgagcgtcag acccgtaga aaagatcaaa ggatcttctt gagatccttt t
ttctgcgc
1681 gtaatctgct gcttgcaaac aaaaaacca ccgtaccag cgtggtttg t
ttgccggat
1741 caagagctac caactctttt tccgaaggta actggcttca gcagagcgca g
ataccaaat
1801 actgtccttc tagtgtagcc gtagttaggg caccacttca agaactctgt a
gcaccgct
1861 acatacctcg ctctgctaact cctgttacca gtgctgctg ccagtggcga t
aagtcgtgt
1921 cttaccgggt tggactcaag acgatagtta ccggataagg cgcagcggtc g
ggctgaacg
1981 gggggttcgt gcacacagcc cagcttgag ogaacgacct acaaccgaact g
agataccca
2041 cagcgtgagc tatgagaaag cgcacgctt cccgaaggga gaaaggcgga c
aggtatcgg
2101 gtaagcggca gggctggaac aggagagcgc acgagggagc ttccaggggg a
aacgcctgg
2161 tatctttata gtctgtcgg gtttcgccac ctctgacttg agcgtcgatt t
ttgtgatgc
2221 tcgtcagggg ggcggagcct atggaaaaac gccagcaacg cggccttttt a
cgttcctg
2281 gccttttctt ggccttttgc tcacatgttc ttctcgtgt tatccctga t
tctgtggat
2341 aaccgtatta ccgcctttga gtgagctgat accgtcgcgc gcagccgaac g
accgagcgc
2401 agcagtcag tgagcgagga agcggaagag cgcctgatgc ggtattttct c
cttacgcat
2461 ctgtcggta ttccacacc catatatggt gcactctcag tacaatctgc t
ctgatgccg
2521 catagttaag ccagtataca ctccgtatc gctacgtgac tgggtcatgg c
tgcgcccgc
2581 acaccgccca acaccgctg acgcgccctg acgggcttgt ctgctccgg c
atccgctta
2641 cagacaagct gtgaccgtct ccgggagctg catgtgtcag aggttttcac c
gtcatcacc

```

FIG 16A (CONT)

WO 02/090300

PCT/US02/14358

```

2701 gaaacgcgcg aggcägcctgc ggtaaagctc atcagcgtgg tcgtgaagcg a
ttoaacagat
2761 gtctgcctgt tcatccgcgt ccagctcgtt gagtttctcc agaagcgta a
tgtctggct
2821 totgataaag cgggccatgt taaggcggtt ttttctctgt ttggtoactg a
tgcctccgt
2881 gtaaggggga tttctgttca tgggggtaat gataccgatg aaacgagaga g
gatgcctcac
2941 gatacgggtt actgatgatg aacatgcccg gttactggaa cgttgtaggg g
taaaccaact
3001 ggcggtatgg atgcggcggg accagagaaa aatcaactcag ggtcaatgcc a
gcgcttctgt
3061 taatacagat gtaggtgttc cacagggtag ccagcagcat cctgcgatgc a
gatccggaa
3121 cataatggtg cagggcgctg acttccgcgt ttccagactt tacgaaacac g
gaaaccgaa
3181 gaccattcat gttgtgtctc aggtcgcaga cgttttgtag cagcagtcgc t
tcacgttcg
3241 ctgcgctatc ggtgattcat totgctaacc agtaaggcaa ccccgccagc c
tagccgggt
3301 cctcaacgac aggagcacga tcatgcgcac ccgtggggcc gccatgccgg c
gataatggc
3361 ctgcttctcg ccgaaacgtt tggtygcggg accagtgacg aaggcttgag c
gaggcgctg
3421 caagattccg aataccgcaa gcgacaggcc gatcatcgtc gcgtccagc g
aaagcggtc
3481 ctgcgcgaaa atgacccaga gcgctgccgg caactgtcct acgagttgca t
gataaagaa
3541 gacagtcata agtgccgga cgatagtcac gccccgcgcc caccggaagg a
gctgactgg
3601 gttgaaggctt ctaaggga toggtcgaga tcccggtgcc taatgagtga g
ctaacttac
3661 attaatggcg ttgcgctcac tgcccgttt ccagtcggga aacctgtcgt g
ccagctgca
3721 ttaatgaatc ggccaacgcg cggggagagg cggtttgct attgggcgcc a
gggtggttt
3781 ttcttttcac cagtgcgacg ggcaacagct gattgccctt caccgcctgg c
cctgagaga
3841 gttgcagcaa gcggtccacg ctggtttgcc ccagcaggcg aaaatcctgt t
tgatggtgg
3901 ttaacggcgg gatataacat gagctgtctt ccgtatcgtc gtatccact a
ccgagatgt
3961 ccgcaccaac gcgcagcccg gactcggtaa tggcacgcat tgcgccagc g
ccatctgat
4021 cgttgccaac cagcatcgca gtgggaacga tgccctcatt cagcatttgc a
tggtttgtt
4081 gaaaaccgga catggcactc cagtcgcctt cccgttcgc tatcggtga a
tttgattgc

```

FIG 16A (CONT)

WO 02/090300

PCT/US02/14358

```

ttaatgggc
4201 ccgctaacag cgcgatttgc tggtagacca atgcgaccag atgctccacg c
ccagtccgg
4261 taccttcttc atgggagaaa ataatactgt tgatgggtgt ctggtcagag a
catcaagaa
4321 ataacgcogg aacattagtg caggcagctt ccacagcaat ggcatcctgg t
catccagg
4381 gatagttaat gatcagccca ctgacgcgtt gcgcgagaag attgtgcacc g
ccgctttac
4441 aggccttcgac gccgcttcgt tctaccatcg acaccaccac gctggcacc c
gttgatcgg
4501 cgcgagattt aatcgccggc acaatttgcg acggcgcgtg cagggccaga c
tggaggtgg
4561 caacgccaat cagcaacgac tgtttgccg ccagttgttg tgccaogcgg t
tgggaatgt
4621 aattcagctc cgcacatgcc gcttccactt tttcccggt tttgcagaa a
cgtggctgg
4681 cctggttcac cagcgggaa acggtctgat aagagacacc ggcatactct g
cgacatcgt
4741 ataacgttac tggtttcaca ttcaccacc tgaattgact ctcttcggg c
gctatcatg
4801 ccataccggc aaaggttttg cgcacatcga tgggtgccg gatctcgacg c
tctccctta
4861 tgcgactcct gcattaggaa gcagcccagt agtaggttga ggccgttgag c
accgcggcc
4921 gcaaggaaatg gtgcatgcaa ggagatggcg cccaacagtc ccccgccac g
gggcctgcc
4981 accataccca cgcgaaaca agcgcctcatg agcccgaaatg ggcgagcccg a
tctcccca
5041 tgggtgatgt cggcgatata ggcgccagca accgcacctg tggcgccggg g
atgcggggc
5101 acgatgcgtc cggcgtagag gatcgagatc gatctcgatc ccgcgaaatt a
atacgactc
5161 actatagggg aattgtgagc ggataacaat tccctctag aaataatttt g
tttaacttt
5221 aagaaggaga tatacatatg aaagaaaccg ctgctgctaa attcgaacgc c
agcacatgg
5281 acagcccaga tctgggtacc ctggtgccac gcggttccat ggotgatato a
gatctaattg
5341 gtctcgaaac tcacaacaca aggcctctgta tctaggaag tggcccagcg g
cacacacgg
5401 cggcgattta cgcagctagg gctgaactta aacctcttct cttcgaagga t
ggatggcta
5461 acgacatcgc tcccggtggc caactaaca ccaaccaga cgtcgagaat t
tccccggat
5521 ttccagaagg tattctcgga gttagagctca ctgacaaatt ccgtaaacaa t
cggagcgat
5581 tgggtactac gatatttaca gagacgggta cgaaagtcca tttctcttcg a

```

FIG 16A (CONT)

WO 02/090300

PCT/US02/14358

aaccgttta
5641 agctattcac agattccaag gccattctcg ctgaogctgt gattctogct a
ctggagctg
5701 tggctaagcg gcttagcttc gttggatctg gtgaagggtc tggagggttc t
ggaaccgtg
5761 gaatctccgc atgcgctgtt tgcgacggag ctgctccgat attccgtaac a
aactctttg
5821 cggatgatcg tggaggcgat tcagcaatgg aagaagcaaa ctttcttaca a
aatatggat
5881 ccaaagtgtataataccat tggatggatg ctttttcggc gtctaagatt a
tgcagcagc
5941 gcgctttgtc taatcctaag attgatgtga ttggaactc gtctgttgtg g
aagcttatg
6001 gagatggaga aagagatgtg cttggaggat tgaagtga gaaatgtggtt a
ccggtgatg
6061 tttctgattt aaaagtttct ggattgttct ttgctattgg tcatgagcca g
ctaccaagt
6121 ttttgatgg tgggtgtgag ttagattcgg atggttatgt tgtaacgaag c
ctggtacta
6181 cacagactag cgttcccggg gttttcgtc cgggtgatgt tcaggataag a
agtataggc
6241 aagccatcac tgcgcagga actgggtgca tggcagcttt ggatgcagag c
attacttac
6301 aagagattgg atctcagcaa ggtaagagtg atggagtcga caagcttgcg g
ccgcac

FIG 16A (CONT)

WO 02/090300

PCT/US02/14358

The WMG variant coding sequence with S-tag at the N-terminus, His-Tag at C-terminus (5238-26)

```

1  tcgagcacca ccaccaccac cactgagatc cggctgctaa caaagccoga a
aggaagctg
61  agttggctgc tgcaccgct gagcaataac tagcataacc ccttggggcc t
ctaaacggg
121  tcttgagggg ttttttgctg aaaggaggaa ctatatccgg attggcgaat g
ggacgcgcc
181  ctgtagcggc gcattaagcg cggcgggtgt ggtgggttacg cgcagcgtga c
cgctacact
241  tgccagcgcc ctacgcgccg ctccctttgc tttcttccct tcctttctcg c
cacgttcgc
301  cggctttccc cgtcaagctc taaatcgggg gctcccttta gggttccgat t
tagtgcttt
361  acggcacctc gaccccaaaa aacttgatta gggtagtgtt tcacgtagt g
gccatcgcc
421  ctgatagacg gtttttcgcc ctttgacgtt ggagtccacg ttctttaata g
tggactctt
481  gttccaaact ggaacaacac tcaacctat ctcggtctat tcttttgatt t
ataagggat
541  tttgccgatt tcggcctatt ggttaaaaaa tgagctgatt taacaaaaat t
taacgcgaa
601  ttttaacaaa atattaacgt ttacaatttc aggtggcact ttcggggaa a
tgtgcgcgg
661  aacccctatt tgtttatttt tetaaataca ttcaaatatg tatccgctca t
gaattaatt
721  cttagaaaaa ctcacgcagc atcaaatgaa actgcaattt attcatatca g
gattatcaa
781  taccatattt ttgaaaaagc cgtttctgta atgaaggaga aaactcacog a
ggcagttcc
841  ataggatggc aagatcctgg tatcgtctcg cgattccgac tcgtccaaca t
caataaac
901  ctattaattt cccctcgta aaaataaggt tatcaagtga gaaatcacca t
gagtgaoga
961  ctgaatccgg tgagaatggc aaaagtttat gcatttcttt ccagacttgt t
caacaggcc
1021  agccattacg ctogtcacga aaatcaactcg catcaaccaa accgttatte a
ttcgtgatt
1081  ggcgcctgagc gagacgaaat acgcgatcgc tggtaaaaag acaattacaa a
caggaatcg
1141  aatgcaaccg gcgcaggaac actgccacg catcaacaat attttcacct g
aatcaggat
1201  attcttctaa tacctggaat gctgttttcc cggggatcgc agtgggtgagt a
accatgcat

```

FIGURE 16B

WO 02/090300

PCT/US02/14358

```

1261 catcaggagt acggataaaa tgcttgatgg tcggaagagg cataaattcc g
tcagccagt
1321 ttagtctgac catctcatct gtaacatcat tggcaacgct acctttgcca t
gtttcagaa
1381 acaactctgg cgcctcgggc ttcccataca atcgatagat tgcgcacact g
attgcccga
1441 cattatcgcg agcccattta taccatata aatcagcacc catgttggaa t
ttaatcgcg
1501 gctagagca agacgtttcc cgttgaatat ggctcataac accccttgta t
tactgttta
1561 tgaagcaga cagttttatt gtccatgacc aaaatccctt aacgtgagtt t
tcgttccac
1621 tgagcgtcag accccgtaga aaagatcaaa ggatcttctt gagatccttt t
tttctgcgc
1681 gtaatctgct gcttgcaaac aaaaaacca ccgctaccag cgggtggtttg t
ttgcggat
1741 caagagctac caactctttt tcggaaggta actgggttca gcagagcgca g
ataccaaat
1801 actgtccttc tagtgtagcc gtagttaggc caccacttca agaactctgt a
gcaccgct
1861 acatacctcg ctctgcta atctgttcca gtggctgctg ccagtggcga t
aagtcgtgt
1921 cttacggggt tggactcaag acgatagtta ccggataagg cgcagcggtc g
ggctgaacg
1981 ggggggttcgt gcacacagcc cagcttggag cgaacgacct acaccgaact g
agataccta
2041 cagcgtgagc tatgagaaag cgcacgctt cccgaaggga gaaaggcgga c
aggatccg
2101 gtaagcggca gggtcggaac aggagagcgc acgagggagc ttccaggggg a
aacgcctgg
2161 tatctttata gtccctgtcgg gtttcgccac ctctgacttg agcgtcgatt t
ttgtgatgc
2221 tcgtcagggg ggcggagcct atggaaaaac gccagcaacg cggccttttt a
cggttcctg
2281 gccttttgct ggccttttgc tcacatgttc tttcctgctg tatcccttga t
tctgtggat
2341 aaccgtatta ccgcctttga gtgagctgat accgctogcc gcagcogaac g
accgagcgc
2401 agcagtcag tgagcgagga agcgggaagag cgcctgatgc ggtattttct c
cttacgcat
2461 ctgtgcggta tttcacaccg catatatggt gcaactctag tacaatctgc t
ctgatgcg
2521 catagttaag ccagtataca ctccgctatc gctacgtgac tgggtcatgg c
tgcccccgc
2581 acacccgcca acaccgctg acgcgcctg acgggcttgt ctgctcccg c
atccgctta
2641 cagacaagct gtgacgtct ccgggagctg catgtgtcag aggttttcac c
gtcatcac

```

FIG 16B (CONT)

WO 02/090300

PCT/US02/14358

2701 gaaacgcgcg aggcagctgc ggtaaagctc atcagcgtgg tcgtgaagcg a
ttcacagat
2761 gtctgcctgt tcatccgcgt ccagctcgtt gagtttctcc agaagcgta a
tgtctgggt
2821 tctgataaag cgggccatgt taaggcggtt tttttcctgt ttggtcactg a
tgctccgt
2881 gtaaggggga tttctgttca tgggggtaat gataccgatg aaacgagaga g
gatgctcac
2941 gatacgggtt actgatgatg aacatgcccg gttactggaa cgttgtgagg g
taaacaaact
3001 ggcggtatgg atgcgcgggg accagagaaa aatcactcag ggtcaatgcc a
gcgcttcgt
3061 taatacagat gtaggtgttc cacagggtag ccagcagcat cctgcgatgc a
gatccggaa
3121 cataatgggt cagggcgctg acttcgcgtt ttccagactt tacgaaacac g
gaaaccgaa
3181 gaccattcat gttgttctc aggtcgcaga cgttttgag cagcagtcgc t
tcacgttcg
3241 ctccgctatc ggtgattcat tctgctaacc agtaaggcaa cccgcgcagc c
tagccgggt
3301 cctcaacgac aggagcacga tcatgcgcac ccgtggggcc gccatgcggc c
gataatggc
3361 ctgottctcg ccgaaacgtt tggtagcggt accagtgacg aaggottgag c
gagggcgtg
3421 caagattccg aataccgcaa gcgacaggcc gatcatcgto gcgctccagc g
aaagcggtc
3481 ctgcgcgaaa atgaccaga gcgctgcgg caccgtgcct acgagttgca t
gataaagaa
3541 gacagtcata agtgcggcga cgtatgtoat gccccgcgc caccggaagg a
gctgactgg
3601 gttgaaggct ctcaaggca tcggtcgaga tcccggtgcc taatgagtga g
ctaacttac
3661 attaatgtcg ttgcgtcac tgcccgctt ccagtcggga aacctgtcgt g
ccagctgca
3721 ttaatgaatc ggccaacgcg cggggagagg cggtttggt attgggcgc a
gggtggtt
3781 ttcttttcac cagtgagacg ggcaacagct gattgccett caccgcctgg c
cotgagaga
3841 gttgcagcaa gcggtccacg ctggtttgcc ccagcaggcg aaaatcctgt t
tgatggtg
3901 ttaacggcgg gatataacat gagctgtctt cggtatcgto gtatccact a
ccgagatgt
3961 ccgcaccaac gcgcagccc gactcggtaa tggcacgcat tgcgccagc g
ccatctgat
4021 cgttggcaac cagcatcgca gtgggaacga tgccctcatt cagcatttgc a
tggtttgtt
4081 gaaaaccgga catggcactc cagtgcctt cccgttcgc tatcggtga a
tttgattgc

FIG 16B (CONT)

WO 02/090300

PCT/US02/14358

```

4141 gaggtagata tttatgccag ccagccagac gcagacgcgc cgagacagaa c
ttaatgggc
4201 ccgctaacag cggatattgc tggtagacca atgcgaccag atgctccacg c
ccagtcgcg
4261 tacogtcttc atgggagaaa ataatactgt tgatgggtgt ctggtcagag a
catcaagaa
4321 ataacgccgg aacattagtg caggcagctt ccacagcaat ggcatcctgg t
catccagcg
4381 gatagttaat gatcagccca ctgacgcgtt gcgcgagaag attgtgcacc g
ccgctttac
4441 aggccttogac gccgcttcgt tctaccatcg acaccaccac gctggcacc a
gttgatcgg
4501 cgcgagattt aatcgccgcg acaatttgcg acggcgcgtg cagggccaga c
tgaggtgg
4561 caacgccaat cagcaacgac tgtttgccg ccagttgttg tgccacgcgg t
tggaatgt
4621 aattcagctc cgcctatgcc gcttcactt tttcccgct tttgcagaa a
cgtggctgg
4681 cctggttcac cagcgggaa acggtctgat aagagacacc ggcatactct g
cgacatcgt
4741 ataacgttac tggtttcaca ttcaccacc tgaattgact ctcttcggg c
gctatcag
4801 ccataccgcg aaaggttttg cgcctatcga tgggtgccg gatctcgacg c
tctccctta
4861 tgcgactcct gcattaggaa gcagcccagt agtaggttga ggcggttag c
accgcgcgc
4921 gcaaggaatg gtgcatgcaa ggagatggcg cccaacagtc ccccgccac g
ggcctgcc
4981 accataccca cgcgaaaca agcgtctatg agccgaagt ggcgagcccg a
tcttccca
5041 tcggtgatgt cggcgatata ggcgcagca accgcacctg tggcgccggt g
atgcgggc
5101 acgatgcgtc cggcgtagag gatcgagatc gatctcgatc ccgcgaaatt a
atacgactc
5161 actatagggg aattgtgagc ggataacaat tccctctag aaataatttt g
tttaacttt
5221 aagaaggaga tatacatatg aaagaaaccg ctgctgctaa attcgaacgc c
agcacatgg
5281 acagcccaga tctgggtacc ctggtgccac gcggttccat ggcctgatac a
gatctaattg
5341 gtctcgaaac tcacaacaca aggcctctgta tcgtaggaag tggcccagcg g
cacacacgg
5401 cggcgattta cgcagctagg gotgaactta aacotcttot ctccaagga t
ggatggcta
5461 acgacatcgc tcccgttgt caactaaca ccaccaccga cgtcgagaat t
tcccggat
5521 ttccagaagg tattctcgga gtagagctca ctgacaaatt ccgtaacaa t
cggagcgat

```

FIG 16B (CONT)

WO 02/090300

PCT/US02/14358

5581 tcggtactac gatatttaca gagacgggtga cgaaagtoga tttctcttcg a
aacggttta
5641 agctattcac agattccaag gccattctcg ctgacgctgt gattctcgct a
ctggagctg
5701 tggctaagcg gcttagcttc gttggatctg gtgaagggtc tggaggtttc t
ggaaccgtg
5761 gaatctccgc atgcgctgtt tgcgacggag ctgctccgat attccgtaac a
aacctcttg
5821 cgggtgatcg tggaggcgat tcagcaatgg aagaagcaaa cttctctaca a
aatatggat
5881 ccaaagtga tataatccat tggatggatg cttttggtgc gtctaagatt a
tgcagcagc
5941 gcgctttgtc taatcctaag attgatgtga ttgggaactc gtctgttgtg g
aagcttatg
6001 gagatggaga aagagatgtg cttggaggat tgaaagtga gaatgtggtt a
ccggtgatg
6061 tttctgattt aaaagtttct ggattgttct ttgctattgg tcctgagcca g
ctaccaagt
6121 ttttggtatg tgggtttgag ttagattcgg atggttatgt tgtcacgaag c
ctggtacta
6181 cacagactag cggtcccgga gttttcgcgt cgggtgatgt tcaggataag a
agtataggc
6241 aagccatcac tgctgcagga actgggtgca tggcagcttt ggatgcagag c
attacttac
6301 aagagattgg atctcagcaa ggtaagagtg atggagtcga caagcttgcg g
ccgcac

FIG 16B (CONT)

WO 02/090300

PCT/US02/14358

The WIS variant coding sequence with S-tag at the N-terminus, His-Tag at C-terminus (5238-26)

```

1  tcgagcacca ccaccaccac cactgagatc cggctgctaa caaagccoga a
aggaagctg
61  agttggctgc tgcaccgct gagcaataac tagcataacc ccttggggcc t
ctaaacggg
121 tcttgagggg tttttgctg aaaggaggaa ctatatccg attggcgaat g
ggaocggcc
181 ctgtagcggc gcattaagcg cggcgggtgt ggtggttacg cgcagcgtga c
cgctacact
241 tgccagcgcc ctagcggccg ctccctttgc tttcttccct tcctttctcg c
cacgttcgc
301 cggctttccc cgtcaagctc taaatcgggg gctcccttta gggttccgat t
tagtgcttt
361 acggcacctc gaccccaaaa aaacttgatta gggtagtggt tcacgtagt g
gcatcgcc
421 ctgatagacg gtttttcgcc ctttgacgtt ggagtccacg ttctttaata g
tggaactctt
481 gttccaaact ggaacaacac tcaaccctat ctcggtctat tcttttgatt t
ataagggat
541 tttgcgatt tcggcctatt ggttaaaaaa tgagctgatt taacaaaaat t
taacgcgaa
601 ttttaacaaa atattaacgt ttacaatttc aggtggcact tttcggggaa a
tgtgcggcg
661 aacccctatt tgtttatttt tctaaataca ttcaaatatg tatccgctca t
gaattaatt
721 cttagaaaaa ctcatcgagc atcaaatgaa actgcaattt attcatatca g
gattatcaa
781 taccatattt ttgaaaagc cgtttctgta atgaaggaga aaactcaccg a
ggcagttcc
841 ataggatggc aagatccctg tatcgggtcg cgattccgac tcgtccaaca t
caatacaac
901 ctattaattt ccctctgtca aaaataaggt tatcaagtga gaaatcacca t
gagtgacga
961 ctgaatccgg tgagaatggc aaaagtttat gcatttcott coagaottgt t
caacaggcc
1021 agccattacg ctgtoatca aaatcactcg catcaaccaa accgttatte a
ttcgtgatt
1081 gcgcctgagc gagacgaaat acgcgatcgc tgttaaaagg acaattacaa a
cagggaatcg
1141 aatgcaaccg gcgcaggaac actgccagcg catcaacaat attttcacct g
aatcaggat
1201 attctttaa tacctggaat gctgttttcc cggggatcgc agtggtgagt a
accatgcat

```

FIGURE 16C

WO 02/090300

PCT/US02/14358

```

1261 catcaggagt acggataaaa tgcttgatgg tcggaagagg cataaattcc g
tcagccagt
1321 ttagtctgac catctcatct gtâacatcat tggcaacgct acctttgcca t
gtttcagaa
1381 acaactctgg cgcctcgggc ttcccataca atcgatagat tgtcgccacct g
attgccga
1441 cattatcgcg agcccattta taccatata aatcagcgc catgttgga t
ttaatcgcg
1501 gcctagagca agacgtttcc cgttgaatat ggctcataac accccttgta t
tactgttta
1561 tgtaagcaga cagttttatt gttcatgacc aaaatccctt aacgtgagtt t
tcgttcac
1621 tgagcgtcag accccgtaga aaagatcaa ggatcttctt gagatccttt t
ttctgcgc
1681 gtaatctgct gottgcaaac aaaaaacca cgcctaccag cgttggtttg t
ttgcggat
1741 caagagctac caactctttt tccgaaggta actggcttca gcagagcgca g
ataccaat
1801 actgtccttc tagttagacc gtagttaggc caccacttca agaactctgt a
gcacgcct
1861 acatacctcg ctctgcta at octgttacca gtgctgctg ccagtggcga t
aagtcgtgt
1921 cttaccgggt tggactcaag acgatagtta ccggataagg cgcagcggtc g
ggctgaacg
1981 ggggggttcgt gcacacagcc cagcttggag cgaacgaact acaccgaact g
agataccta
2041 cagcgtgagc tatgagaaag cggcacgctt cccgaaggga gaaaggcgga c
aggatccg
2101 gtaagcggca gggtcggaac aggagagcgc acgagggagc ttccaggggg a
aacgcctgg
2161 tatctttata gtctgtcgg gtttcgccac ctctgacttg agcgtcgatt t
ttgtgatgc
2221 tcgtcagggg ggcggagcct atggaaaaac gccagcaacg cggccttttt a
cggttcctg
2281 gccttttgct ggccttttgc tcacatgttc ttctctgctt tatccctga t
totgtggat
2341 aaccgtatta ccgcctttga gtgagctgat accgctcgcc gcagccgaac g
accgagcgc
2401 agcagtcag tgagcgagga agcgggaagag cgctctgatc ggtattttct c
cttacgcat
2461 ctgtgcggta tttcacacog catatatggt gcaactctcag tacaatctgc t
ctgatgcgc
2521 catagttaag ccagtataca ctccgctatc gctacgtgac tgggtcatgg c
tgcccccgc
2581 acaccgcga acaccgcgt acgcgcctg acgggcttgt ctgctcccgc c
atccgctta
2641 cagacaagct gtgaccgtct ccgggagctg catgtgtcag aggtttttac c
gtcatcac

```

FIG 16C (CONT)

WO 02/090300

PCT/US02/14358

2701 gaaacgcgcg aggcagctgc ggtaaagctc atcagcgtgg tcgtgaagcg a
 ttacacagat
 2761 gtctgcctgt tcctccgcgt ccagctcgtt gagtttctcc agaagcgta a
 tgtctggct
 2821 tctgataaag cgggccatgt taagggcggt ttttctctgt ttggtcaactg a
 tgcctccgt
 2881 gtaaggggga tttctgttca tgggggtaat gataccgatg aaacgagaga g
 gatgctcac
 2941 gatacgggtt actgatgatg aacatgcccg gttactggaa cgttgtgagg g
 taaacaact
 3001 ggcggtatgg atgcggcggg accagagaaa aatcactcag ggtcaatgcc a
 gcgcttcgt
 3061 taatacagat gtaggtgttc cacagggtag ccagcagcat cctgcgatgc a
 gatccggaa
 3121 cataatggtg cagggcgctg acttcgcgtt ttccagactt taagaaacac g
 gaaaccgaa
 3181 gaccattcat gttgttgctc aggtcgaga cgttttgag cagcagtcgc t
 tcacgttcg
 3241 ctgcggtato ggtgattcat tctgctaacc agtaaggcaa ccccgccagc c
 tagccgggt
 3301 cctcaacgac aggagcacga tcatgcgcac cgtgggggcc gccatgccgg c
 gataatggc
 3361 ctgcttctcg ccgaaacgtt tggtaggggg accagtgacg aaggcttgag c
 gaggggctg
 3421 caagattccg aataccgcaa gcgacagccc gatcatcgte gcgctccagc g
 aaagcggtc
 3481 ctgcgccgaa atgaccaga gcgctgcggg cactgtcct acgagttgca t
 gataaagaa
 3541 gacagtcata agtgccggcga cgaatgcat gcccgcgcg caccggaagg a
 gctgactgg
 3601 gttgaaggct ctcaagggca tcggtcgaga tcccggtgca taatgagtga g
 ctaacttac
 3661 attaatgctg ttgcgtcac tgcccgcttt ccagtcggga aacctgtcgt g
 ccagctgca
 3721 ttaatgaato ggcoaacgcg cggggagagg cggtttgctt attgggcgac a
 gggtaggtt
 3781 ttcttttcac cagtgcagcg ggcaacagct gattgocctt caccgcttg c
 cctgagaga
 3841 gttgcagcaa ggggtccagc ctggtttgcc ccagcaggcg aaaatcctgt t
 tgatggtag
 3901 ttaacggcgg gatataacat gagctgtctt cggtatcgte gtatccact a
 ccgagatgt
 3961 ccgcaccaac ggcagcccg gactcggtta tggcaagcat tgcgccagc g
 ccactctgat
 4021 cgttggaac cagcatcgca gtgggaacga tgccctcatt cagcatttgc a
 tggtttgtt
 4081 gaaaaccgga catggcactc cagtgcctt cccgttcgc tatcggtga a
 tttgattgc

FIG 16C (CONT)

WO 02/090300

PCT/US02/14358

```

4141 gagtgagata tttatgccag ccagccagac gcagacgcgc cgagacagaa c
ttaatgggc
4201 ccgctaacag cgcgatttgc tggtagacca atgcgaccag atgctccacg c
ccagtcgcg
4261 taccgctcttc atgggagaaa ataatactgt tgatgggtgt ctggtcagag a
catcaagaa
4321 ataacgcgcg aacattagtg caggcagctt ccacagcaat ggcatacttg t
catccagcg
4381 gatagttaat gatcagccca ctgacgcgtt gcgcgagaag attgtgcacc g
ccgctttac
4441 aggccttcgac gccgcttcgt tctaccatcg acaccaccaac gctggcaccc a
gttgatcgg
4501 cgcgagattt aatcgccgcg acaatttgcg acggcgcggtg caggggcaga c
tggaggttg
4561 caacgccaat cagcaacgac tgtttgcccg ccagttgttg tgccacgcgg t
tgggaatgt
4621 aattcagctc cgccatcgcc gcttccactt tttcccgctt ttctgcagaa a
cgtggtcgg
4681 cctggttcac cagcggggaa acggtctgat aagagacacc ggcatactct g
cgacatcgt
4741 ataacgttac tggtttcaca ttcaccaccc tgaattgact ctcttcggg c
gctatcatg
4801 ccataccgcg aaaggttttg cgccattcga tgggtgcgg gatctcgacg c
tctccctta
4861 tgcgactcct gcattaggaa gcagcccagt agtaggttga ggccgttgag c
accgcccgc
4921 gcaagggaatg gtgcattcaa ggagatggcg cccaacagtc ccccgccac g
gggcctgccc
4981 accataccca cgcgaaaaca agcgtctatg agcccgaaat ggcgagcccg a
tcttcccca
5041 tcggtgatgt cggcgatata ggcgcoagca accgcacctg tggcgccggt g
atgcgggcc
5101 acgatgcgtc cggcgtagag gatcgagatc gatctcgatc cgcgaaatt a
atacgactc
5161 actatagggg aattgtgagc ggataacaat tccctctag aaataatttt g
tttaacttt
5221 aagaaggaga tatacatatg aaagaaaccg ctgctgctaa attogaacgc c
agcacatgg
5281 acagcccaga tctgggtacc ctggtgccac gcggttccat ggtgatato a
gatctaatt
5341 gtctcgaaac tcacaacaca aggtctgtga togtaggaag tggcccacg g
cacacagg
5401 cggcgattta cgcagctagg gctgaactta aacctcttct cttoagaagg t
ggatggcta
5461 acgacatgc tcccgggtgt caactaaca caccacoga cgtogagaat t
tcccggat
5521 ttccagaagg tattctcga gtagagctca ctgacaaatt ccgtaacaa t
cgagcgat

```

FIG 16C (CONT)

WO 02/090300

PCT/US02/14358

5581 tcggtactac gatatttaca gagacggtga cgaaagtoga tttctotteg a
aaccgttta
5641 agctattcac agattccaag gccattctcg ctgacgctgt gattctcgct a
ctggagctg
5701 tggctaagcg gcttagcttc gtggatctg gtgaagggtc tggagggttc t
ggaaccgtg
5761 gaatctccgc atgcgctgtt tgcgacggag ctgctcogat attccgtaac a
aacctcttg
5821 cggtgatcgg tggaggcgat tcagcaatgg aagaagcaaa cttctttaca a
aatatggat
5881 ccaaagtgtg tataatccat tggattgatg cttttctgc gtctaagatt a
tgcagcagc
5941 gcgctttgtc taatcotaag attgatgtga tttggaactc gtctgttggtg g
aagcttatg
6001 gagatggaga aagagatgtg ctggaggat tgaagtga gaatgtggtt a
ccggtgatg
6061 tttctgattt aaaagttctt ggattgttct ttgctattgg tcatgagcca g
ctaccaagt
6121 ttttggtggtg tgggtgtgag ttagattcgg atggttatgt tgtcacgaag c
ctgggtacta
6181 cacagactag cgttcccga gttttcgctg cgggtgatgt tcaggataag a
agtataggc
6241 aagccatcac tgcgcagga actgggtgca tggcagcttt ggatgcagag c
attacttac
6301 aagagattgg atctcagcaa ggtaagagt atggagtcga caagcttgcg g
ccgcac

FIG 16C (CONT)

WO 02/090300

PCT/US02/14358

The WMS variant coding sequence with S-tag at the N-terminus, His-Tag at C-terminus (5238-26)

```

1  tcgagcacca ccaccaccac cactgagatc cggctgctaa caaagcccga a
aggaagctg
61  agttggctgc tgccaccgct gagcaataac tagcataacc ccttggggcc t
ctaaacggg
121  tcttgagggg ttttttgctg aaaggaggaa ctatatccgg attggcgaat g
ggacgcgcc
181  ctgtagcggc gcattaagcg cggcgggtgt ggtggttacg cgcagcgtga c
cgctacact
241  tgccagcgcc ctacgcgccg ctccctttgc tttcttccct tcctttctcg c
cacgttcgc
301  cggctttccc cgtcaagctc taaatcgggg gctcccttta gggttccgat t
tagtgcttt
361  acggcacctc gaccccaaaa aacttgatta gggatggtt tcacgtagt g
gccatcgcc
421  ctgatagacg gtttttcgcc ctttgacgtt ggagtccacg ttctttaata g
tggactctt
481  gttccaaact ggaacaacac tcaaccctat ctcggtctat tcttttgatt t
ataagggat
541  tttgccgatt tcggcctatt ggtaaaaaa tgagctgatt taacaaaaat t
taacgcgaa
601  ttttaacaaa atattaacgt ttacaatttc aggtggcact tttcggggaa a
tgtgcgcgg
661  aacccctatt tgtttatttt tctaaataca ttcaaatatg tatccgctca t
gaattaatt
721  cttagaaaaa ctcatcgagc atcaaatgaa actgcaattt attcatatca g
gattatcaa
781  taccatatct ttgaaaaagc cgtttctgta atgaaggaga aaactcaccg a
ggcagttcc
841  ataggatggc aagatcctgg tatcgggtctg cgattccgac tcgtccaaca t
caatacaac
901  ctattaattt cccctcgtca aaaataaggt tatcaagtga gaaatcacca t
gagtgcgga
961  ctgaatccgg tgagaatggc aaaagtattt gcatttcttt ccagacttgt t
caacaggcc
1021  agccattacg ctcgatcatc aaatcactcg catcaaccaa accgttatcc a
ttcgtgatt
1081  gcgcctgagc gagacgaaat acgcgatcgc tgtaaaagg acaattacaa a
caggaatcg
1141  aatgcaaccg gcgcaggaac actgccagcg catcaacaat attttcacct g
aatcaggat
1201  attcttctaa tacctggaat gctgttttcc cggggatcgc agtggtgagt a
accatgcat
1261  catcaggagt acggataaaa tgcttgatgg tcggaagagg cataaattcc g

```

FIGURE 16D

40/113

WO 02/090300

PCT/US02/14358

```

tcagccagt
1321 ttagtctgac catctcatct gtaacatcat tggcaacgct acctttgcca t
gtttcagaa
1381 acaactctgg cgcacgaggc ttcccatata atcgatagat tgcgcacac t
attgccga
1441 cattatcgcg agcccattta taccatata aatcagcaco catgttgga t
ttaatcgcg
1501 gcctagagca agacgtttcc cgttgaatat ggctcataac accccttgta t
tactgttta
1561 tgtaagcaga cagttttatt gttcatgacc aaaatccett aacgtgagtt t
tcgttccac
1621 tgagcgtcag acccgtaga aaagatcaaa ggatcttctt gagatccttt t
ttctcgcg
1681 gtaatctgct gcttgcaaac aaaaaaacca ccgctaccag cgggtggttg t
ttgcggat
1741 caagagctac caactctttt tccgaaggta actggcttca gcagagcgca g
ataccaaat
1801 actgtccttc tagttagacc gtagttaggc caccacttca agaactctgt a
gcacgcct
1861 acatacctcg ctctgctaatt cctgttaacca gtggtgctg cagtgaggca t
aagtcgtgt
1921 cttaccgggt tggactcaag acgatagtta ccgataagg cgcagcggtc g
ggctgaacg
1981 gggggttcgt gcacacagcc cagcttgagg cgaacgacct acaccgaact g
agataccta
2041 cagcgtgagc tatgagaaag cgcacgctt cccgaaggga gaaaggcgga c
aggtatccg
2101 gtaagcgcca gggtcggaac aggagagcgc acgagggagc ttccaggggg a
aacgcctgg
2161 tatctttata gtccgtcgg gtttcgccac ctctgacttg agcgtcgatt t
ttgtgatgc
2221 tcgtcagggg ggccggagcct atggaaaaac gccagcaacg cggccttttt a
cggttcctg
2281 gccttttgcg ggccctttgc tcacatgttc tttcctgctg tatccctga t
tctgtggat
2341 aaccgtatta ccgcctttga gtgagctgat accgctcgcc gcagccgaac g
accgagcgc
2401 agcaggtcag tgagcgagga agcgggaagag cgcctgatgc ggtattttct c
cttacgcat
2461 ctgtgcggta ttccacaccg catatatggt gcaactctcag tacaatctgc t
ctgatgccg
2521 catagttaag ccagtataca ctccgctatc gctacgtgac tgggtcatgg c
tgcccccg
2581 acaccgccca acaccgctg acgcgcctg acgggcttgt ctgctccgg c
atccgctta
2641 cagacaagct gtgaccgtct cggggagctg catgtgtcag aggttttcac c
gtcatcac
2701 gaaacgcgag aggcagctgc ggtaaagctc atcagcgtgg tcgtgaagcg a

```

FIG 16D (CONT)

WO 02/090300

PCT/US02/14358

```

ttcacagat
2761 gtctgcctgt tcatccgcgt ccagctcggt gagtttctcc agaagcgta a
tgtctgggt
2821 tctgataaag cgggccatgt taaggcggt ttttctctgt ttggtcactg a
tgcctccgt
2881 gtaaggggga tttctgttca tgggggtaat gataccgatg aaacgagaga g
gatgctcac
2941 gatacgggtt actgatgatg aacatgcccg gttactggaa cgttgtgagg g
taaacaaact
3001 ggcggtatgg atgcggcggg accagagaaa aatcaactcag ggtaaatgcc a
gcgcttcgt
3061 taatacagat gtaggtgttc cacagggtag ccagcagcat cctgcgatgc a
gatccggaa
3121 cataatgggtg cagggcgctg acttcgcgtt ttccagactt tacgaaacac g
gaaacgaa
3181 gaccattcat gttgttctc aggtcgaga cgttttgag cagcagtcgc t
tcacgttcg
3241 ctgcgttato ggtgattcat tctgtaacc agtaaggcaa cccgcgcagc c
tagccgggt
3301 cctcaacgac aggagcacga tcatgcgcac ccgtggggcc gccatgccgg c
gataatggc
3361 ctgcttctcg ccgaaacgtt tggtgccggg accagtgacg aaggcttgag c
gaggcgctg
3421 caagattccg aataccgcaa gcgacagccc gatcatcgtc gcgctccagc g
aaagcggtc
3481 ctgccgaaa atgaccaga gcgctgcgg cactgtcct accagttgca t
gataaagaa
3541 gacagtcata agtgcggcga ccatagtcac gccccgcgcc cacgggaagg a
gctgactgg
3601 gttgaaggct ctcaagggca tcggtcgaga tcccggtgcc taatgagtga g
ctaaattac
3661 attaatgctg ttgcgctcac tgcccgctt ccagtcggga aacctgtcgt g
ccagctgca
3721 ttaatgaate ggccaacgcg cggggagagg cggtttgctg attgggcgcc a
gggtgggtt
3781 ttcttttcae cagtgcgacg ggcaacagct gattgccctt caccgcctgg c
cctgagaga
3841 gttgcagcaa gcggtccacg ctggtttgcc ccagcaggcg aaaatcctgt t
tgatggtgg
3901 ttaacggcgg gatataacat gagctgtctt cgtatcgtc gtatccact a
ccgagatgt
3961 ccgcaccaac gcgcagcccg gactcggtaa tggcaagcat tgcgccagc g
ccatctgat
4021 cgttggaac cagcatcgca gtgggaadga tgcctcatt cagcatttgc a
tggtttggt
4081 gaaaaccgga catggcactc cagtcgcctt cccgttcgc tatcggtga a
tttgattgc
4141 gagtgcgata tttatgccag ccagccagac gcagacgcgc cgagacagaa c

```

FIG 16D (CONT)

WO 02/090300

PCT/US02/14358

```

4141 gagtgagata tttatgccag ccagccagac gcagacgcgc cgagacagaa c
ttaatgggc
4201 ccgctaacag cgcgatttgc tggtagacca atgcgaccag atgctccacg c
ccagtgcgc
4261 taccgtcttc atgggagaaa ataatactgt tgatgggtgt ctggtcagag a
catcaagaa
4321 ataacgcgcg aacattagtg caggcagctt ccacagcaat ggcacccctgg t
catccagcg
4381 gatagttaat gatcagccca ctgacgcggt gcgcgagaag attgtgcacc g
ccgctttac
4441 aggcttcgac gccgcttctg tctaccatcg acaccaccac gctggcacc c
gttgatcgc
4501 cgcgagattt aatcgccgcg acaatttgcg acggcgcgtg cagggccaga c
tggaggtgg
4561 caacgccaat cagcaacgac tgtttgcccg ccagtgttg tgcacgcgcg t
tgggaatgt
4621 aattcagctc cgcctatcgc gcttccactt ttcccgctt ttccgcagaa a
cgtggctgg
4681 cctggttoac cagcgggaa acggtctgat aagagacacc ggcatactct g
cgacatcgt
4741 ataacgttac tggtttcaca ttcaccacc tgaattgact ctcttcggg c
gctatcatg
4801 ccataccgcg aaaggttttg cgccttcga tgggttcgg gatctcgag c
tctccctta
4861 tgcgactcct gcattaggaa gcagcccagt agtaggttga ggcggttgag c
accgcgcgc
4921 gcaagggaatg gtgcagcaa ggagatggcg ccacaacagt ccccgccac g
gggcctggc
4981 accataccca cgcgaaaca agcgtcatg agcccgaagt ggcgagccg a
tcttcccca
5041 tcggtgatgt cgcgatata ggcgcagca accgcacctg tggcgccggg g
atgccgggc
5101 acgatgcgtc cggcgtagag gatcgagac gatctcgac ccgcgaaatt a
atacgactc
5161 actatagggg aattgtgagc ggataacaat tccctctag aaataatttt g
tttaacttt
5221 aagaaggaga tatacatatg aaagaaaccg ctgctgctaa attcgaacgc c
agcacatgg
5281 acagcccaga tctgggtacc ctggtgccac gcggttccat ggctgatato a
gatctaattg
5341 gtctcgaaac tcacaacaca aggtctgtga tcgtaggaag tggcccagcg g
cacacacgg
5401 cggcgattta cgcagctagg gctgaactta aacctctct ctctgaagga t
ggatggcta
5461 acgacatcgc tcccgtgggt caactaaca ccaccaccga cgtcgagaat t
tccccggat
5521 ttocagaagg tattctcgga gtagagctca ctgacaaatt ccgtaacaa t
cggagcgat

```

FIG 16D (CONT)

WO 02/090300

PCT/US02/14358

5581 tcggtactac gatatttaca gagacggtga cgaaagtcga tttctcttcg a
aaccgttta
5641 agctattcac agattccaag gccattctcg ctgacgtgt gattctcgt a
ctggagctg
5701 tggctaagcg gcttagcttc gttggatctg gtgaaggttc tggaggttc t
ggaaccgtg
5761 gaatctccgc atgcgtgtt tgcgacggag ctgctccgat attccgtaac a
aacctcttg
5821 cggatgacgg tggagggcat tcagcaatgg aagaagcaaa ctttcttaca a
aatatggat
5881 ccaaagtgtg tataatccat tgggtggatg cttttcgggc gtctaagatt a
tgcagcagc
5941 gcgctttgtc taatccaaag attgatgtga tttggaactc gtctgtttgtg g
aagcttatg
6001 gagatggaga aagagatgtg cttggaggat tgaagtgaag gaatgtggtt a
ccggtgatg
6061 tttctgattt aaaagtttct ggattgttct ttgotattgg tcatgagcca g
ctaccaaagt
6121 ttttggatgg tgggttgag ttagattcgg atggttatgt tgtcagaag c
ctggtacta
6181 cacagactag cgttcccgga gtttctcgtg cgggtgatgt tcaggataag a
agtataggc
6241 aagccatcac tgcgcagga actgggtgca tggcagcttt ggatgcagag c
attacttac
6301 aagagattgg atctcagcaa ggtaagagt atggagtcca caagcttgcg g
ccgcac

FIG 16D (CONT)

WO 02/090300

PCT/US02/14358

The WLS variant coding sequence with S-tag at the N-terminus, His-Tag at C-terminus (5238-26)

```

1  tcgagcacca ccaccaccac cactgagatc cggctgctaa caaagcccga a
aggaagctg
61  agttggctgc tgccaccgct gagcaataac tagcataacc ccttggggcc t
ctaaacggg
121  tcttgagggg ttttttctg aaaggaggaa ctatatccgg attggcgaat g
ggacgcggcc
181  ctgtagcggc gcattaagcg cggcgggtgt ggtggttaag cgcagcgtga c
cgctacact
241  tgccagcgcc ctgagcccg ctcttttcgc tttcttcctt tcctttctcg c
cacgttcgc
301  cggctttccc cgtcaagctc taaatcgggg gctcccttta gggttccgat t
tagtgcttt
361  acggcacctc gaccccaaaa aacttgatta gggatggtt tcacgtagt g
gccatcgcc
421  ctgatagacg gttttctgcc ctttgacgtt ggagtccacg ttctttaata g
tggactctt
481  gttccaaact ggaacaacac tcaaccctat ctcggtctat tcttttgatt t
ataagggat
541  tttgcgattt tcggcctatt ggttaaaaaa tgagctgatt taacaaaaat t
taacgcgaa
601  ttttaacaaa atattaacgt ttacaatttc aggtggcact ttcggggaa a
tgtgcgcgg
661  aacccctatt tgtttatttt tctaaataca ttcaaatacg tatccgctca t
gaattaatt
721  cttagaaaaa ctcatcgagc atcaaatgaa actgcaattt attcatatca g
gattatcaa
781  taccatattt ttgaaaaagc cgtttctgta atgaaggaga aaactcacog a
ggcagttcc
841  ataggatggc aagatcctgg tatcgggtctg cgaattccgac tcgtocaaca t
caatacaac
901  ctattaattt cccctcgtca aaaataaggt tatcaagtga gaaatcacca t
gagtgcga
961  ctgaatccgg tgagaatggc aaaagtttat gcatttcttt ccagacttgt t
caacaggcc
1021  agccattacg ctcgatcatc aaatcactcg catcaaccac accgttattc a
ttcgtgatt
1081  gcgcctgagc gagacgaaat acgcgatcgc tgttaaaagg acaattacaa a
caggaatcg
1141  aatgcaaccg gcgcaggaa actgccagcg catcaacaat attttcacct g
aatcaggat
1201  attcttctaa tacctggaat gctgttttcc cggggatcgc agtgggtgagt a
accatgcat

```

FIGURE 16E

WO 02/090300

PCT/US02/14358

1261 catcaggagt acggataaaa tgcttgatgg tcggaagagg cataaatcc g
 tcagccagt
 1321 ttagtctgac catctcatct gtaacatcat tggcaacgct acctttgcca t
 gtttcagaa
 1381 acaactctgg cgcacgggc ttcccatata atcgatagat tgtcgcacct g
 attgcccga
 1441 cattatcgg agccattta taccatata aatcagcgc catgttgaa t
 ttaatcgg
 1501 gcctagagca agcgtttcc cgttgaatat ggctcataac accccttgta t
 tactgttta
 1561 tgtaagcaga cagttttatt gttcatgacc aaaatccctt aacgtgagtt t
 tcgttccac
 1621 tgagcgtcag acccgtaga aaagatcaaa ggatcttctt gagatccttt t
 tttctgcg
 1681 gtaatctgct gcttgcaaac aaaaaacca cgcctaccag cgggtggtttg t
 ttgcggat
 1741 caagagctac caactctttt tccgaagga actggcttca gcagagcgca g
 ataccaat
 1801 actgtcttc tagtgtagcc gtggttaggc caccacttca agaactctgt a
 gcacgcct
 1861 acatacctcg ctctgctaact cctgttacca gtggctgctg ccagtggcga t
 aagtcgtgt
 1921 cttaccgggt tggactcaag acgatagtta ccggataagg cgcagcggtc g
 ggctgaacg
 1981 ggggggtctg gcaacagcc cagcttgag cgaacgacct acaccgaact g
 agataccta
 2041 cagcgtgagc tatgagaaag ccgcacgctt cccgaaggga gaaaggcgga c
 aggtatccg
 2101 gtaagcgca gggtcggaac aggagagcgc acgagggagc ttccaggggg a
 aacgcctgg
 2161 tatctttata gtctgtcgg gtttcgccac ctctgacttg agcgtcgatt t
 ttgtgatgc
 2221 tcgtcagggg ggcggagcct atggaaaaac gccagcaacg cggccttttt a
 cggttcctg
 2281 gcccttttgc ggccttttgc tcacatgttc ttctctcgt tatccctga t
 tctgtgat
 2341 aaccgtatta ccgccttga gtgagctgat accgtcgcgc gcagccgaac g
 accgagcgc
 2401 agcagagcag tgagcgagga agcgggaagag cgcctgatgc ggtattttct c
 cttacgcat
 2461 ctgtgcggta ttccacacc catatatggt gcaactctcag tacaatctgc t
 ctgatgcg
 2521 catagttaag ccagtataca ctccgctatc gctacgtgac tgggtcatgg c
 tgccccc
 2581 acaccgccca acaccgctg acgcgcctg acgggcttgt ctgctcccg c
 atccgctta
 2641 cagacaagct gtgaccgtct ccgggagctg catgtgtcag aggtttttac c
 gtcacacc

FIG 16E (CONT)

WO 02/090300

PCT/US02/14358

```

2701 gaaacgcgcg aggcagctgc ggtaaagctc atcagcgtgg tcgtgaagcg a
ttcacagat
2761 gtctgcctgt tcatccgctt ccagctcgtt gagttttctc agaagcgta a
tgtctggct
2821 tctgataaag cgggccatgt taaggcggtt tttttcctgt ttggtcactg a
tgcctccgt
2881 gtaaggggga tttctgttca tgggggtaat gataccgatg aaacgagaga g
gatgctcac
2941 gatacgggtt actgatgatg aacatgcccg gttactggaa cgttgtgagg g
taaacaaact
3001 ggcggtatgg atgcggcggg accagagaaa aatcactcag ggtcaatgcc a
gcgttctgt
3061 taatacagat gtaggtgttc cacagggtag ccagcagcat cctgcgatgc a
gatccggaa
3121 cataatgggt cagggcgctg acttccgctt ttccagactt tacgaaacac g
gaaaccgaa
3181 gaccattcat gttgttgctc aggtgcgaga cgttttgtag cagcagtcgc t
tcacgttctg
3241 ctgcgctatc ggtgattcat tctgctaacc agtaaggcaa ccccgccagc c
tagccgggt
3301 cctcaacgac aggagcacga tcatgcgcac ccgtggggcc gccatgccgg c
gataatggc
3361 ctgcttctcg ccgaaacgtt tgggtggcggg accagtgcag aaggcttgag c
gagggcgctg
3421 caagattccg aataccgcaa gcgacaggcc gatcatcgtc gcgctccagc g
aaagcggtc
3481 ctgcgcgaaa atgaccaga gcgctgccgg caoctgtcct acgagttgca t
gataaagaa
3541 gacagtcata agtgccggca cgatagtcac gcccgcgccc caccggaagg a
gctgactgg
3601 gttgaaggct ctcaagggca tcggtcgaga tcccggtgcc taatgagtga g
ctaacctac
3661 attaatgctg ttgcgctcac tgcccgtttt ccagtcggga aacotgtcgt g
ccagctgca
3721 ttaatgaatc ggccaacgcg cggggagagg cggtttgctt attgggcgcc a
gggtgggtt
3781 ttcttttcac cagtgcgacg ggcaacagct gattgccctt caccgcctgg c
cctgagaga
3841 gttgcagcaa gcggtccacg ctggtttgcc ccagcaggcg aaaatcctgt t
tgatgggtg
3901 ttaacggcgg gatataacat gagctgtctt cggtatcgtc gtatcccaact a
ccgagatgt
3961 ccgcaccaac gcgcagcccg gactcggtaa tggcaacgat tgcgcccagc g
ccatctgat
4021 cgttggcaac cagcatcgca gtgggaacga tgccctcatt cagcatttgc a
tggtttgtt
4081 gaaaacggga catggcactc cagtcgcctt cccgttcgcg tatcgggtga a
tttgattgc

```

FIG 16E (CONT)

WO 02/090300

PCT/US02/14358

```

4141 gagtgagata tttatgccag ccagccagac gcagacgcgc cgagacagaa c
ttaatgggc
4201 ccgctaacag cgcgatttgc tggtagacca atgcgaccag atgctccacg c
ccagtcgcg
4261 taccgctcttc atgggagaaa ataatactgt tgatgggtgt ctggtcagag a
catcaagaa
4321 ataacgccgg aacattagtg caggcagctt ccacagcaat ggcaccttg t
catccagcg
4381 gatagttaat gatcagccca ctgacgcgtt gcgcgagaag attgtgcacc g
ccgctttac
4441 aggccttcgac gccgcttcgt tctaccatcg acaccaccac gctggcaccc a
gttgatcgg
4501 cgcgagattt aatcgccgcg acaatttgcg acgcgcgctg cagggccaga c
tggaggtgg
4561 caacgccaat cagcaacgac tgtttgcccg ccagttgttg tgccacgcgg t
tgggaatgt
4621 aattcagctc cgccatcgcc gcttccactt tttccgcgt ttcgcagaa a
cgtggctgg
4681 cctggttcac cgcgcgggaa acggtctgat aagagacacc ggcatactct g
cgacatcgt
4741 ataacgttac tggtttcaca ttcaccaccc tgaattgact ctcttcggg c
gctatcatg
4801 ccataccgcg aaagggtttg cgccattcga tgggtgctcg gatctcgacg c
tctccctta
4861 tgcgactcct gcattaggaa gcagcccagt agtaggttga ggcggttgag c
accgcccgc
4921 gcaaggaaatg gtgcatgcaa ggagatggcg cccaacagtc ccccgccac g
gggcctgcc
4981 acoataccca cgcgaaaca agcgcctcatg agcccgaagt ggcgagccc a
tcttccca
5041 tcggtgatgt cggcgatata ggcccgca accgcacctg tggcgccggt g
atgcggcc
5101 acgatgcgtc cggcgtagag gatcgagatc gatctcgatc ccgcgaaatt a
atacgactc
5161 actatagggg aattgtgagc ggataacaat tcccctctag aaataatttt g
ttaacttt
5221 aagaaggaga tatacatatg aaagaacccg ctgctgctaa attcgaacgc c
agcacatgg
5281 acagcccaga tctgggtacc ctggtgccac gcggttccat ggctgatatc a
gatctaag
5341 gtctcgaaac tcacaacaca aggctctgta tcgtaggaag tggcccagcg g
cacacacgg
5401 cggcgattta cgcagctagg gctgaactta aacctcttct cttcgaagga t
ggatggcta
5461 acgacatcgc tcccggtggc caactaaca ccaccaccga cgtcgagaat t
tcccggat
5521 ttccagaagg tattctcgga gtagagctca ctgacaaatt ccgtaaacaa t
cggagcgat

```

FIG 16E (CONT)

48/113

WO 02/090300

PCT/US02/14358

5581 tcggtactac gatatttaca gagacgggtga cgaaagtoga tttctcttcg a
aaccgttta
5641 agctattcac agattccaag gccattctcg ctgacgctgt gattctcgct a
ctggagctg
5701 tggctaagcg gcttagcttc gttggatctg gtgaagggtc tggagggttc t
ggaaccgtg
5761 gaatctccgc atgcgctgtt tgogacggag ctgctccgat attccgtaac a
aacctcttg
5821 cgggtgatcgg tggaggcgat tcagcaatgg aagaagcaa ctttcttaca a
aatatggat
5881 ccaaagtgtataataccat tgggtggatg cttttctgc gtctaagatt a
tgacgacgc
5941 gcgctttgtc taatcctaag attgatgtga ttggaactc gtctgttgtg g
aagcttatg
6001 gagatggaga aagagatgtg cttggagatg tgaagtga gaatgtggtt a
ccggtgatg
6061 tttctgattt aaaagtttct ggattgttct ttgctattgg tcatgagcca g
ctaccaagt
6121 ttttggatgg tgggtgtgag ttagattcgg atggttatgt tgtcacgaag c
ctggtacta
6181 cacagactag cgttcccgga gttttcgtg cgggtgatgt tcaggataag a
agtataggc
6241 aagccatcac tgctgcagga actgggtgca tggcagcttt ggatgcagag c
attacttac
6301 aagagattgg atctcagcaa ggtaagagtg atggagtoga caagcttgcg g
ccgcac

FIG 16E (CONT)

WO 02/090300

PCT/US02/14358

The WRT variant coding sequence with S-tag at the N-terminus, His-Tag at C-terminus (5238-26)

```

1  tcgagcacca caccaccac cactgagatc cggctgctaa caaagccoga a
aggaagctg
61  agttggctgc tgcacccgct gagcaataac tagcataacc cctggggcc t
ctaaacggg
121  tcttgagggg tttttgctg aaaggaggaa ctatatccgg attggcgaat g
ggacgcgcc
181  ctgtagcggc gcattaagcg cggcgggtgt ggtgggtacg cgcagcgtga c
cgctacact
241  tgcacagcgc ctagcgcccg ctcttttcgc tttcttccct tcctttctcg c
caggttcgc
301  cggcttttcc cgtcaagctc taaatcgggg gctcccttta gggttcgat t
tagtgcttt
361  acggcacctc gaccccaaaa aacttgatta ggtgatggt tcacgtagt g
gccatcgcc
421  ctgatagacg gtttttcgcc ctttgacgtt ggagtccacg ttctttaata g
tggactctt
481  gttccaaact ggaacaacac tcaacctat ctcggtctat tcttttgatt t
ataagggat
541  ttggccgatt tcggcctatt ggttaaaaaa tgagctgatt taacaaaaat t
taacgcgaa
601  ttttaacaaa atattaacgt ttacaatttc aggtggcact ttcgggggaa a
tgtgcgcgg
661  aaccctctat tgtttatttt tctaaatata ttcaaataatg tatccgctca t
gaattaat
721  cttagaaaaa ctcacgcgag atcaaatgaa actgcaattt attcatatca g
gattatcaa
781  taccatattt ttgaaaaagc cgtttctgta atgaaggaga aaactcacgc a
ggcagttcc
841  ataggatggc aagatcctg tatcggtctg cgattccgac tcgtccaaca t
caatacaac
901  ctattaattt cccctcgtca aaaataaggt tatcaagtga gaaatcacca t
gagtgacga
961  ctgaatccgg tgagaatggc aaaagtttat gcatttcttt ccagacttgt t
caacaggcc
1021  agccattacg ctcgatcatc aaatcactcg catcaaccaa accgttatto a
ttcgtgatt
1081  gcgcctgagc gagacgaaat acgcgatcgc tgttaaaagg acaattacaa a
caggaatcg
1141  aatgcaaccg gcgcagggaac actgccagcg catcaacaat attttcacct g
aatcaggat
1201  attcttctaa tacctggaat gctgttttcc cgggggatcgc agtgggtgagt a
accatgcat

```

FIGURE 16F

WO 02/090300

PCT/US02/14358

```

1261 catcaggagt acggataaaa tgcttgatgg tcggaagagg cataaattcc g
tcagccagt
1321 ttagtctgac catctcatct gtaacatcat tggcaacgct acctttgcca t
gtttcagaa
1381 acaactctgg cgcacgggc tccccatata atcgatagat tgcgcacct g
attgccga
1441 cattatcgcg agccatttta taccoatata aatcagcctc catgttggaa t
ttaatcgcg
1501 gcctagagca agacgtttcc cgttgaatat ggctcataac accccttgta t
tactgttta
1561 tgtaagcaga cagttttatt gttcatgacc aaaatccctt aacgtgagtt t
tcgttccac
1621 tgagcgtcag acccogtaga aaagatcaaa ggatcttctt gagatccttt t
ttctgogc
1681 gtaatctgct gcttgcaaac aaaaaaacca ccgctaccag cggcgggttg t
ttgcggat
1741 caagagctac caactctttt tcggaaggta actggcttca gcagagcgca g
ataccaaat
1801 actgtccttc tagtgtagcc gtagttaggg caccacttca agaactotgt a
gcacgcct
1861 acatacctcg ctctgctaatt cctgttacca gtggtctgtg ccagtggcga t
aagtctgtg
1921 cttaccgggt tggactcaag acgatagtta ccggataagg cgcagcggtc g
ggctgaacg
1981 gggggttcgt gcacacagcc cagcttgagg cgaacgcct acaccgaact g
agataccta
2041 cagcgtgagc tatgagaaag cgcacgcctt ccggaaggga gaaaggcgga c
aggatatcg
2101 gtaagcggca gggtcggaac aggagagcgc acgagggagc ttccaggggg a
aacgcctgg
2161 tatctttata gtctgtcgg gtttcgccac ctctgacttg agcgtcgatt t
ttgtgatgc
2221 tcgtcagggg ggcggagcct atggaaaaac gccagcaacg cggccttttt a
cggttcctg
2281 gccttttgct ggccttttgc tcacatgttc tttcctgctg tatccctga t
ctgtggat
2341 aacggtatta ccgcctttga gtgagctgat accgctcgcc gcagccgaac g
accgagcgc
2401 agcagtcag tgagcgagga agcgggaagag cgcctgatgc ggtatttct c
cttaogcat
2461 ctgtcgggta ttccacacg catatatggt gcactctcag tacaatctgc t
ctgatgccc
2521 catagttaag ccagtataca ctccgctatc gctacgtgac tgggtcatgg c
tgcgcccgc
2581 acaccccgca acacccgctg acgcgccctg acgggcttgt ctgctcccg c
atccgctta
2641 cagacaagct gtgaccgtct ccgggagctg catgtgtcag aggttttcac c
gtcatcacc

```

FIG 16F (CONT)

WO 02/090300

PCT/US02/14358

2701 gaaacgcgcg aggcagctgc ggtaaagctc atcagcgtgg tcgtgaagcg a
ttcacagat
2761 gtctgcctgt tcatccgcgt ccagctcgtt gattttctcc agaagcgtta a
tgtctggct
2821 tctgataaag cgggccatgt taaggccggt tttttcctgt ttggtcaactg a
tgccctcgt
2881 gtaaggggga tttctgttca tgggggtaat gataccgatg aaacgagaga g
gatgctcac
2941 gatacggggt actgatgatg aacatgcccg gttactggaa cgttgtgagg g
taaaccaact
3001 ggcggtatgg atgcggcggg accagagaaa aatcactcag ggtcaatgcc a
gcgcttcgt
3061 taatacagat gtaggtgttc cacagggtag ccagcagcat cctgcgatgc a
gatccggaa
3121 cataatgggt cagggcgctg acttccgcgt ttccagactt tacgaaacac g
gaaacgaa
3181 gaccattcat gttgttctc aggtcgcaga cgttttgag cagcagtcgc t
tcacgttcg
3241 ctgcgtatc ggtgattcat tctgctaacc agtaaggcaa ccccgccagc c
tagccgggt
3301 cctcaacgac aggagcacga tcatgcgcac ccgtggggcc gccatgccgg c
gataatggc
3361 ctgcttctcg ccgaaacgtt tgggtggcgg accagtgaag aaggcttgag c
gagggcggtg
3421 caagattccg aalaccgcaa gcgacaggcc gatcatcgtc gcgctccagc g
aaagcggtc
3481 ctgcgcgaaa atgaccaga gcgctgccg cacctgtcct acgagttgca t
gataaagaa
3541 gacagtcata agtgcggcga cgatagtcac gccccgcgcc caccggaagg a
gctgactgg
3601 gttgaaggct ctcaagggca tcggtcgaga tcccggtgcc taatgagtga g
ctaacttac
3661 attaatggcg ttgcgctcac tgcccgttt ccagtcggga aacctgtcgt g
ccagctgca
3721 ttaatgaatc ggccaacgcg cggggagagg cggtttgct attgggcgcc a
gggtgggtt
3781 ttcttttcac cagtgcagc ggcaacagct gattgccctt caccgcctgg c
cctgagaga
3841 gttgcagcaa gcggtccacg ctggtttgcc ccagcaggcg aaaatcctgt t
tgatgttg
3901 ttaacggcgg gatataacat gagctgtctt cggtatcgtc gtatccact a
ccgagatgt
3961 ccgcaccaac gcgcagccc gactcggtaa tggcaagcat tgcgccagc g
ccatctgat
4021 cgttggcaac cagcatcgca gtgggaacga tgcctcatt cagcatttgc a
tggtttgtt
4081 gaaaaaccga catggcactc cagtcgcctt cccgttccgc tatcgggtga a
tttgattgc

FIG 16F (CONT)

WO 02/090300

PCT/US02/14358

```

4141 gagttagata tttatgccag ccagccagac gcagacgcgc cgagacagaa c
ttaatgggc
4201 ccgctaacag cgcgatttgc tggtagacca atgogaccag atgctccacg c
ccagtcgcg
4261 taccgtcttc atgggagaaa ataatactgt tgatgggtgt ctggtcagag a
catcaagaa
4321 ataacgccgg aacattagtg caggcagctt ccacagcaat ggcatcctgg t
catccagcg
4381 gatagttaat gatcagccca ctgacgcgtt gcgcgagaag attgtgcacc g
cgcctttac
4441 aggccttcgac gccgcttcgt tctaccatcg acaccaccac gctggcacc c
gttgatcgg
4501 cgcgagattt aatcgccgcg acaatttgcg acggcgcggt caggggccaga c
tggagggtgg
4561 caacgccaat cagcaacgac tgtttgcccg ccagttgttg tgccacgcgg t
tgggaatgt
4621 aattcagctc cgcctatgcc gcttccactt ttcccgcggt ttccgcagaa a
cgtggctgg
4681 cctggttcac cgcgcgggaa acggtctgat aagagacacc ggcatactct g
cgacatcgt
4741 ataacgttac tggtttcaca ttcaccacc cgaattgact ctcttcggg c
gctatcatg
4801 ccataccgcg aaaggttttg cgccattcga tgggtgtccg gatctcgacg c
tctccctta
4861 tgcgactcct gcattaggaa gcagcccagt agtaggttga ggccgttgag c
accgcgcgc
4921 gcaaggaatg gtgcatgcaa ggagatggcg cccaacagtc ccccgccac g
gggcctgcc
4981 accataccca cgcgaaaca agcgtctatg agcccgaaat ggcgagccc a
tcttcccca
5041 tcggtgatgt cggcgatata ggcgccagca accgcacctg tggcgccggt g
atgccggcc
5101 acgatgcgtc cggcgtagag gatcgagac gatctcgac ccgcgaaatt a
atacgactc
5161 actatagggg aattgtgagc ggataacaat tcccctctag aaataatttt g
tttaacttt
5221 aagaaggaga tatacatatg aaagaaacgc ctgctgctaa attcgaaacg c
agcacatgg
5281 acagcccaga tctgggtacc ctggtgccac gcggttccat ggctgatatc a
gatctaattg
5341 gtctcgaaac tcacaacaca aggctctgta tcgtaggaag tggcccagcg g
cacacagg
5401 cggcgattta cgcagctagg gctgaactta aacctcttct cttcgaagga t
ggatggcta
5461 acgacatcgc tcccgggtgt caactaaca ccaccaccga cgtcgagaat t
tcccggat
5521 ttccagaagg tattctcgga gtagagctca ctgacaaatt ccgtaacaa t
cgagcgat

```

FIG 16F (CONT)

WO 02/090300

PCT/US02/14358

```
5581 tcggtactac gatatttaca gagacggtga cgaaagtcga tttctottog a
aacggttta
5641 agctattcac agattccaag gccattctcg ctgacgctgt gattctcgct a
ctggagctg
5701 tggctaagcg gcttagcttc gttggatctg gtgaagggtc tggaggtttc t
ggaacggtg
5761 gaatctccgc atgcgctgtt tgcgacggag ctgctccgat attccgtaac a
aacctottg
5821 cggatgacgg tggaggcgat tcagcaatgg aagaagcaaa ctttcttaca a
aatatggat
5881 ccaaagtgtg tataatccat tggcgtgatg cttttactgc gtctaagatt a
tgcagcagc
5941 gcgctttgtc taatcctaag attgatgtga tttggaactc gtctgttggtg g
aagcttatg
6001 gagatggaga aagagatgtg cttggaggat tgaagtgaa gaatgtggtt a
ccggtgatg
6061 tttctgattt aaaagtttct ggattgttct ttgctattgg tcatgagcca g
ctaccaagt
6121 ttttggatgg tgggtgttgag ttagattcgg atggttatgt tgtcacgaag c
ctggtacta
6181 cacagactag cgttcccga gttttcgctg cgggtgatgt tcaggataag a
agtataggc
6241 aagccatcac tgctgcagga actgggtgca tggcagcttt ggatgcagag c
attacttac
6301 aagagattgg atctcagcaa ggtaagagtg atggagtcga caagcttgcg g
ccgcac
```

FIG 16F (CONT)

WO 02/090300

PCT/US02/14358

The RYN variant coding sequence with S-tag at the N-terminus, His-Tag at C-terminus (5238-26)

```

1  tcgagcacca ccaccaccac cactgagatc cggtctgtaa caaagcccga a
aggaagctg
61  agttggctgc tgccacogct gagcaataac tagcataacc ccttggggcc t
ctaaacggg
121  tcttgagggg ttttttctg aaaggaggaa ctatatccgg attggcgaat g
ggacgcggc
181  ctgtagcggc gcattaagcg cggcgggtgt ggtggttacg cgcagcgtga c
cgctacact
241  tgccagcgcc ctacgcgcc ctcccttcgc tttcttcctc tcctttctcg c
cacgttcgc
301  cggctttccc cgtcaagctc taaatcgggg gctcccttta gggttccgat t
tagtgcttt
361  acggcacctc gaccccaaaa aacttgatta gggatggtg tcacgtagt g
gccatcgcc
421  ctgatagacg gtttttcgcc ctttgacgtt ggagtccacg ttctttaata g
tggactctt
481  gttccaaact ggaacaacac tcaaccctat ctgggtctat tcttttgatt t
ataagggat
541  tttgcccatt toggcctatt ggtaaaaaa tgagctgatt taacaaaaat t
taacgcgaa
601  ttttaacaaa atattaacgt ttacaatttc aggtggcact tttcggggaa a
tgtgcgcgg
661  aaccctattt tgtttatttt tctaaataca ttcaaatacg tatccgctca t
gaatttaatt
721  cttagaaaaa ctcatcgagc atcaaatgaa actgcaattt attcatatca g
gattatcaa
781  taccatattt ttgaaaaagc cgtttctgta atgaaggaga aaactcaccc a
ggcagttcc
841  ataggatggc aagatcctgg tatcggctcg cgattccgac tcgtccaaca t
caatacaac
901  ctattaattt ccctcgtca aaaataaggt tatcaagtga gaaatcacca t
gagtgaacga
961  ctgaatccgg tgagaatggc aaaagtttat gcatttcttt ccagacttgt t
caacaggcc
1021  agccattacg ctcgctatca aaatcactcg catcaaccaa accgttattc a
ttcgtgatt
1081  gcgcctgagc gagacgaaat acgcgatcgc tgttaaaagg acaattacaa a
caggaatcg
1141  aatgcaaccg gcgcaggaaac actgccagcg catcaacaat attttcacct g
aatcaggat
1201  attcttctaa tacctggaat gctgttttcc cggggatcgc agtggtagt a
accatgcat
1261  catcaggagt acggataaaa tgcttgatgg tcggaagagg cataaattcc g

```

FIGURE 16G

WO 02/090300

PCT/US02/14358

```

tcagccagt
1321 ttagttgtgac catctcatct gtaacatcat tggcaacgct acctttgcca t
gtttcagaa
1381 acaactctgg cgcacggggc ttcccatata atcgatagat tgcgcacct g
attgcccga
1441 cattatcggc agcccattta taccatata aatcagcatc catgttgga t
ttaatcggc
1501 gcctagagca agacgtttcc cgttgaatat ggtcataac accccttgta t
tactgttta
1561 tgtaagcaga cagttttatt gttcatgacc aaaatccctt aacgtgagtt t
tcgttccac
1621 tgagcgtcag accccgtaga aaagatcaaa ggatcttctt gagatccctt t
tttctggc
1681 gtaattgtct gcttgcaaac aaaaaaacca ccgctaccag cgggtggttg t
ttgcccgat
1741 caagagctac caactctttt tccgaaggta actggcttca gcagagcgca g
ataccaaat
1801 actgtccttc tagttagacc gtagttaggc caccacttca agaactctgt a
gcaccgcct
1861 acatacctcg ctctgcta at cctgttacca gtggctgctg ccagtggcga t
aagtcgtgt
1921 cttaccgggt tggactcaag acgatagtta ccggataagg cgcagcggtc g
ggctgaacg
1981 ggggggtctg gcacacagcc cagcttgag cgaacgacct acacogaact g
agataccta
2041 cagcgtgagc tatgagaaag cgcacgctt ccgaaggga gaaaggcggga c
aggtatccg
2101 gtaagcggca gggtcggaac aggagagcgc acgagggagc ttccaggggg a
aacgcctgg
2161 tatctttata gtctgtcgg gtttcgcaac ctctgacttg agcgtcgatt t
ttgtgatgc
2221 tcgtcagggg ggcggagcct atggaaaaac gccagcaacg cggccttttt a
cggttctg
2281 gccttttgc ggccttttgc tcacatgttc tttcctgctg tatccctga t
tctgtgat
2341 aaccgtatta ccgcctttga gtgagctgat accgctcgcc gcagccgaac g
accgagcgc
2401 agcgagtcag tgagcgagga agcggaagag cgcctgatgc ggtatctct c
cttacgcat
2461 ctgtgcggta tttcacaccg catatatggt gcactctcag tacaatctgc t
ctgatgccg
2521 catagttaag ccagtataca ctccgctatc gctacgtgac tgggtcatgg c
tgcccccg
2581 acaccgccca acaccgctg acgcgccctg acgggcttgt ctgctcccg c
atccgctta
2641 cagacaagct gtgacgtct ccgggagctg catgtgtcag aggttttcac c
gtcatcacc
2701 gaaacggcgg aggcagctgc ggtaagctc atcagcgtgg tegtgaagc a

```

FIG 16G (CONT)

WO 02/090300

PCT/US02/14358

ttcacagat
2761 gtctgcctgt tcatccgcgt ccagctcggt gagtttctcc agaagcggtta a
tgtctggct
2821 tctgataaag cgggccatgt taaggcggt tttttcctgt ttggtcactg a
tgcctccgt
2881 gtaaggggga tttctgttca tgggggtaat gataccgatg aaacgagaga g
gatgctcac
2941 gatacgggtt actgatgatg aacatgcccg gttactggaa cgttgtgagg g
taaacaact
3001 ggcggtatgg atgcggcggg accagagaaa aatcactcag ggtcaatgcc a
gcgcttcgt
3061 taatacagat gtaggtgttc cacagggtag ccagcagcat cctgcgatgc a
gatccggaa
3121 cataatgggt cagggcgctg acttccgctg ttccagactt tacgaaacac g
gaaaccgaa
3181 gaccattcat gttgttgctc aggtcgcaga cgttttgag cagcagtcgc t
tcacgttcg
3241 ctgcgctatc ggtgattcat tctgctaacc agtaaggcaa ccccgccagc c
tagccgggt
3301 cctcaacgac aggagcacga tcatgcgcac ccgtggggcc gccatgccgg c
gataatggc
3361 ctgcttctcg ccgaaacgtt tgggtggcgg accagtgacg aaggcttgag c
gagggcggtg
3421 caagattccg aataccgcaa gcgacaggcc gatcatcgtc gcgtccagc g
aaagcggtc
3481 ctgcgcgaaa atgaccaga gcgtgcggg cactgtctct acgagttgca t
gataaagaa
3541 gacagtcata agtcggcgca cgatagtcac gccccgcgcc caccggaagg a
gctgactgg
3601 gttgaaggct ctcaagggca tcggtcgaga tcccgtgcc taatgagtga g
ctaacttac
3661 attaatggcg ttgcgtcac tgcccgcttt ccagtcggga aacctgtcgt g
ccagctgca
3721 ttaatgaatc ggccaacgcg cggggagagg cggtttgctg attgggcgcc a
gggtgggtt
3781 ttctttttcac cagtgcgac ggcaacagct gattgccctt caccgcctgg c
cctgagaga
3841 gttgcagcaa gcggtccacg ctggtttgcc ccagcaggcg aaaatcctgt t
tgatgggtg
3901 ttaacggcgg gatataacat gagctgtctt cggtatcgtc gtatccact a
ccgagatgt
3961 ccgcaccaac gcgcagcccg gactcggtaa tggcacgcat tgcgcccagc g
ccatctgat
4021 cgttggcaac cagcatcgca gtgggaacga tgcctcatt cagcatttgc a
tggtttgtt
4081 gaaaaccgga catggcactc cagtcgcctt cccgttcgcg tatcggtga a
tttgattgc
4141 gagtgcagata tttatgccag ccagccagac gcagacgcgc cgagacagaa c

FIG 16G (CONT)

WO 02/090300

PCT/US02/14358

```

ttaatgggc
4201 cgcctaacag cgcgatttgc tggtagacca atgcgaccag atgctccacg c
ccagtcgcg
4261 tacctgtctc atgggagaaa ataatactgt tgatgggtgt ctggtcagag a
catcaagaa
4321 ataacgccgg aacattagtg caggcagctt ccacagcaat ggcctcctgg t
catccagcg
4381 gatagttaat gatcagccca ctgacgcgtt gcgcgagaag attgtgcacc g
cgcctttac
4441 aggccttcgac gccgcttcgt tctaccatcg acaccaccac gctggcacc a
gttgatcgg
4501 cgcgagattt aatcgccgag acaatttgcg acggcgcgtg cagggccaga c
tggaggtgg
4561 caacgccaat cagcaacgac tgtttgccg ccagttgttg tgccacgcgg t
tgggaatgt
4621 aattcagctc cgcctcgcgc gcttccactt ttcccgctg ttctgcagaa a
cgtggctgg
4681 cctggttcac cagcgggaa acggtctgat aagagacacc ggcatactct g
cgacatcgt
4741 ataacgttac tggtttaca ttcaccacc tgaattgact ctcttcggg c
gctatcatg
4801 ccataccgag aaaggttttg cgccttcga tgggtgtccg gatctcgacg c
tctcctta
4861 tgcgactcct gcattaggaa gcagcccgat agtaggttga ggcggttgag c
accgcggcc
4921 gcaaggaatg gtgcattcaa ggagatggcg ccaacagtc ccccgccac g
gggcctgcc
4981 accataccca cgcgaacaa agcgcctcatg agcccgaggt ggcgagccg a
tcttccca
5041 tccgtgatgt cggcgatata ggcgccagca accgcacctg tggcgccggt g
atgcgggcc
5101 acgatgcgtc cggcgtagag gatcgagatc gatctcgatc ccgcgaaatt a
atacgactc
5161 actatagggg aattgtgagc ggataacaat tccctctag aaataatttt g
tttaacttt
5221 aagaaggaga tatacatatg aaagaaaccg ctgctgctaa attcgaacgc c
agcacatgg
5281 acagcccaga tctgggtacc ctggtgccac gcggttccat ggtgatata a
gatctaag
5341 gtctcgaaac tcacaacaa aggcctctgta tctaggaag tggcccgag g
cacacacgg
5401 cggcgattta cgcagctagg gctgaactta aacctctct ctctgaagga t
ggatggcta
5461 acgacatgc tccgggtggt caactaaca ccaccaccga cgtcgagaat t
tcccgcat
5521 ttccagaagg tattctcgga gtagagctca ctgacaaatt ccgtaacaa t
cggagcgat
5581 tccgtactac gatatttaca gagacggtga cgaagtcga ttctctctg a

```

FIG 16G (CONT)

WO 02/090300

PCT/US02/14358

aaccgttta
5641 agctattcac agattccaag gccattctcg ctgacgctgt gattctcgct a
ctggagctg
5701 tggctaagcg gcttagcttc gttggatctg gtgaaggctc tggaggcttc t
ggaaccgtg
5761 gaatctccgc atgcgctggt tgcgacggag ctgctccgat attccgtaac a
aacctcttg
5821 cggatgacgg tggaggcgat tcagcaatgg aagaagcaaa ctttcttaca a
aatatggat
5881 ccaaagtgtg tataatccat cgctacgatg cttttaacgc gtctaagatt a
tgcagcagc
5941 gcgctttgtc taatcctaag attgatgtga ttggaactc gtctgttggtg g
aagcttatg
6001 gagatggaga aagagatgtg cttggaggat tgaaagtga gaatgtggtt a
ccggtgatg
6061 tttctgattt aaaagtttct ggattgttct ttgctattgg tcatgagcca g
ctaccaagt
6121 ttttggatgg tgggtttgag ttagattcgg atggttatgt tgtcacgaag c
ctggacta
6181 cacagactag cgttcccgga gttttcgctg cgggtgatgt tcaggataag a
agtataggc
6241 aagccatcac tgcgcagga actgggtgca tggcagcttt ggaagcagag c
attacttac
6301 aagagattgg atctcagcaa ggtaagagtg atggagtcga caagcttgcg g
ccgcac

FIG 16G (CONT)

WO 02/090300

PCT/US02/14358

The RFN-A variant coding sequence with S-tag at the N-terminus, His-Tag at C-terminus (5238-26)

```

1  tcgagcacca ccaccaccac cactgagatc cggctgctaa caaagcccca a
aggaagctg
61  agttggctgc tgccaccgct gagcaataac tagcataacc ccttggggcc t
ctaaacggg
121  tcttgagggg ttttttctg aaaggaggaa ctatatccgg attggcgaat g
ggacgcggcc
181  ctgtagcggc gcattaagcg cggcgggtgt ggtggttacg cgcagcgtga c
cgctacact
241  tgccagcgcc ctacgcgccg ctcccttcgc tttcttccct tctttctcg c
cacgttcgc
301  cggctttccc cgtcaagctc taaatcgggg gctcccttta gggttccgat t
tagtgcttt
361  acggcacctc gaccccaaaa aacttgatta gggtagtggt tcacgtagt g
gccatcgcc
421  ctgatagacg gtttttcgcc ctttgacgtt ggagtccacg ttctttaata g
tggactctt
481  gttccaaact ggaacaacac tcaaccctat ctoggtctat tcttttgatt t
ataagggat
541  tttgccgatt tcggcctatt ggttaaaaaa tgagctgatt taacaaaaat t
taacgcgaa
601  ttttaacaaa atattaacgt ttacaatttc agtgggcact tttcggggaa a
tgtgcgcgg
661  aaccctctatt tgtttatttt tctaaatata ttcaaatatg tatccgctca t
gaattaatt
721  cttagaaaaa ctcatcgagc atcaaatgaa actgcaattt attcatatca g
gattatcaa
781  taccatattt ttgaaaaagc cgtttctgta atgaaggaga aaactcacg a
ggcagttcc
841  ataggatggc aagatcctgg tatcggtctg cgattccgac tcgtccaaca t
caatacaac
901  ctattaattt cccctcgta aaaataaggt tatcaagtga gaaatcacca t
gagtgacga
961  ctgaatccgg tgagaatggc aaaagtatat gcatttcttt ccagacttgt t
caacaggcc
1021  agccattacg ctgctcatca aaatcactcg catcaaccaa accgttatto a
ttcgtgatt
1081  ggcctgagc gagacgaat acgcgatcgc tgttaaaagg acaattacaa a
caggaatcg
1141  aatgcaacgg gcgcaggaac actgccagcg catcaacaat attttcacct g
aatcaggat
1201  attcttctaa tacctggaat gctgttttcc cggggatcgc agtggtgagt a
accatgcat

```

FIGURE 16H

WO 02/090300

PCT/US02/14358

1261 catcaggagt acggataaaa tgcttgatgg tcggaagagg cataaattoc g
tcagccagt
1321 ttagtctgac catctcatct gtaacatcat tggcaacgct acctttgccca t
gtttcagaa
1381 acaactctgg cgcacgggc ttcccatata atcgatagat tgctgcacct g
attgcccga
1441 cattatcgcg agcccattta taccatata aatcagcatc catgttggaa t
ttaatcgcg
1501 gcctagagca agacgtttcc cgttgaatat ggctcataac accccttgta t
tactgttta
1561 tgtaagcaga cagttttatt gttcatgacc aaaatocctt aacgtgagtt t
tcgttccac
1621 tgagcgtag accccgtaga aaagatcaaa ggatcttctt gagatccttt t
tttctgcg
1681 gtaatctgct gcttgcaaac aaaaaaacca ccgctaccag cggtggtttg t
ttgcccgat
1741 caagagctac caactctttt tccgaagga actggcttca gcagagcgca g
ataccaaat
1801 actgtccttc tagttagacc gtagttagga caccacttca agaactctgt a
gcaccgct
1861 acatacctcg ctctgcta at cctgttacca gtggtgctg cagtgggcga t
aagtcgtgt
1921 cttaccgggt tggaactcaag acgatagta ccggataagg cgcagcggtc g
ggctgaacg
1981 ggggggttcgt gcacacagcc cagcttgga cgaacgacct acaccgaact g
agatacota
2041 cagcgtgagc tatgagaaaag cgcacgctt cccgaaggga gaaaggcgga c
aggtatccg
2101 gtaagcggca gggcgggaac aggagagcgc acgagggagc ttccaggggg a
aacgcctgg
2161 tatctttata gtctgtcgg gtttcgccac ctctgacttg agcgtcgatt t
ttgtgatgc
2221 tcgtcagggg ggcggagcct atggaaaaac gccagcaacg cggccttttt a
cggttcctg
2281 gccttttct ggccttttgc tcacatgttc ttctctgctg tatccctga t
tctgtggat
2341 aaccgtatta ccgcctttga gtgagctgat accgctcgcc gcagccgaac g
accgagcgc
2401 agcaggtcag tgagcgagga agcgggaagag cgcctgatgc ggtattttct c
cttacgcat
2461 ctgtgcggta ttccacaccg catatatggt gcactctcag tacaatctgc t
ctgatgccg
2521 catagttaag ccagtataca ctccgctatc gctacgtgac tgggtcatgg c
tgcccccgc
2581 acaccgcga acaccgctg acgcgccctg acgggcttgt ctgctcccgg c
atccgctta
2641 cagacaagct gtgaccgtct ccgggagctg catgtgtcag aggttttca c
gtcatcacc

FIG 16H (CONT)

61/113

WO 02/090300

PCT/US02/14358

```

2701 gaaacgcgcg aggcagctgc ggtaaagctc atcagcgtgg tcgtgaagcg a
ttcacagat
2761 gtctgcctgt tcattccgct ccagctcgtt gagttttctcc agaagcgta a
tgtctggct
2821 tctgataaag cgggccatgt taaggcggtt ttttccctgt ttggtcactg a
tgccctcgt
2881 gtaaggggga tttctgttca tgggggtaat gataccgatg aaacgagaga g
gatgctcac
2941 gatacgggtt actgatgatg aacatgcocg gttactggaa cgttgtaggg g
taaacact
3001 ggcggtatgg atgcggcggg accagagaaa aatcactcag ggtcaatgcc a
gcgcttcgt
3061 taatacagat gtaggtgttc cacagggtag ccagcagcat cctgcgatgc a
gatccggaa
3121 cataatgggt cagggcgctg acttccgctt ttccagactt tacgaaacac g
gaaacgaa
3181 gaaccattcat gttgttgctc aggtgcgaga cgtttttcag cagcagtcgc t
tcacgttcg
3241 ctgcgctatc ggtgattcat tctgctaacc agtaaggcaa ccccgccagc c
tagccgggt
3301 cctcaacgac agggagcaga tcatgcgcac ccgtggggcc gccatgccgg c
gataatggc
3361 ctgcttctcg ccgaaacgtt tggtaggggg accagtgcag aaggtttgag c
gagggcggt
3421 caagattccg aataccgcaa gcgacaggcc gatcatcgtc gcgctccagc g
aaagcggtc
3481 ctgcgccaaa atgaccgaga gcgctgcggg cactgtctct acgagttgca t
gataaagaa
3541 gacagtcata agtgccgaga cgatagtcac gccccgcgcc caccggaagg a
gctgactgg
3601 gttgaaggct ctcaaggcca tcggtcgaga tcccggtgcc taatgagtga g
ctaacttac
3661 attaatggcg ttgcgctcac tgcccgcttt ccagtcggga aacctgtcgt g
ccagctgca
3721 ttaatgaatc ggccaacgcg cggggagagg cggtttgctt attgggcgcc a
gggtggttt
3781 ttcttttcac cagtgcgacg ggcaacagct gattgocctt caccgcctgg c
cctgagaga
3841 gttgcagcaa gcggtccacg ctggtttgcc ccagcaggcg aaaatcctgt t
tgatggtgg
3901 ttaacggcgg gatataacat gagctgtctt cgggtatcgtc gtatcccact a
ccgagatgt
3961 ccgcaccaac gcgcagcccg gactcggtaa tggcagcat tgcgccagc g
ccatctgat
4021 cgttggaac cagcatcgca gtgggaacga tgccctcatt cagcatttgc a
tggtttgtt
4081 gaaaaaccga catggcactc cagtcgcctt cccgttccgc tatcggctga a
tttgattgc

```

FIG 16H (CONT)

WO 02/090300

PCT/US02/14358

```

4141 gagttagata tttatgccag ccagccagac gcagacgcgc cgagacagaa c
ttaatgggc
4201 ccgctaacag cgcgatttgc tggtagacca atgcgaccag atgctccacg c
ccagtcgcg
4261 taccgtcttc atgggagaaa ataatactgt tgatgggtgt ctggtcagag a
catcaagaa
4321 ataacgcggg aacattatgt caggcagctt ccacagcaat ggcacccctg t
catccagcg
4381 gatagttaat gatcagccca ctgacgcgtt gcgcgagaag attgtgcacc g
ccgctttac
4441 aggtcttgac gccgcttcgt tctaccatcg acaccaccac gctggcacc c
gttgatcgg
4501 cgcgagattt aatgcgcgcg acaatttgcg acggcgcgtg cagggccaga c
tggaggtgg
4561 caacgccaat cagcaacgac tgtttgcgcg ccagttgttg tgccacgcgg t
tgggaatgt
4621 aattcagctc cgcctatgcc gcttccactt tttccgcgt tttcgcagaa a
cgtggctgg
4681 cctggtttac cagcgggaa acggtctgat aagagacacc ggcatactct g
cgacatcgt
4741 ataacgttac tggtttcaca ttcaccacc tgaattgact ctcttcggg c
gctatcatg
4801 ccataccgcg aaaggttttg cgcattcga tgggtccgg gatctcgacg c
tctccctta
4861 tgcgactcct gcattagaa gcagccagc agtaggttga ggcggttgag c
accgcgcgc
4921 gcaaggaatg gtgcatgcaa ggagatggcg cccaacagtc ccccgccac g
gggcctggc
4981 accataccca cgcgaaaca agcgtcatg agcccgaaat ggcgagccg a
tcttccca
5041 tcggtgatgt cggcgatata ggcgcagca accgcacctg tggcgccggg g
atgcgcggc
5101 acgatgcgtc cggcgtagag gatcgagatc gatctcgatc ccgcgaaatt a
atacgactc
5161 actatagggg aattgtgagc ggataacaat tccctctag aaataatttt g
tttaacttt
5221 aagaaggaga tatacatatg aaagaaacg ctgctgctaa attcgaacgc c
agcacatgg
5281 acagccaga tctgggtacc ctggtgccac gcggttccat ggtgatatc a
gatctaatt
5341 gtctcgaac tcacaacaca aggtctgtta tcgtaggaag tggccacgcg g
cacacacgg
5401 cggcgattta cgcagctagg gotgaactta aacctcttct cttcgaagga t
ggatggcta
5461 acgacatgc tcccgtggt caactaaca ccaccacga cgtcgagaat t
tcccggat
5521 ttcagaagg tattctcgga gttagctca ctgacaaatt ccgtaacaa t
cggagcgat

```

FIG 16H (CONT)

WO 02/090300

PCT/US02/14358

5581 tcggtactac gatatttaca gagacggtga cgaaagtoga tttctcttgc a
aaccgttta
5641 agctattoac agattccaag gccattctgc ctgacgtgt gattctgcct a
ctggagctg
5701 tggctaagcg gcttagcttc gttggatctg gtgaaggttc tggaggtttc t
ggaaccgtg
5761 gaatctccgc atgcgctgtt tgcgacggag ctgctccgat attccgtaac a
aaccctttg
5821 cggatgacgg tggaggcgat tcagcaatgg aagaagcaaa ctttcttaca a
aatatggat
5881 ccaaagtgtg tataatccat cgctttgatg cttttaacgc ggctaagatt a
tgcagcagc
5941 ggcctttgtc taatcctaag attgatgtga tttggaactc gtctgttgtg g
aagcttatg
6001 gagatggaga aagagatgtg cttggaggat tgaagtgaa gaatgtggtt a
ccggtgatg
6061 tttctgattt aaaagtttct ggattgttct ttgctattgg tcatgagcca g
ctaccaagt
6121 ttttgatgg tgggtgtgag ttagattcgg atggttatgt tgtcacgaag c
ctggtacta
6181 cacagactag cgttcccga gtttctgcctg cgggtgatgt tcaggataag a
agtataggc
6241 aagccatcac tgcgcagga actgggtgca tggcagcttt ggatgcagag c
attacttac
6301 aagagattgg atctcagcaa ggtaagagtg atggagtcca caagcttgcg g
ccgcac

FIG 16H (CONT)

WO 02/090300

PCT/US02/14358

The RFN variant coding sequence with S-tag at the N-terminus, His-Tag at C-terminus (5238-26)

```

1  tcgagcacca ccaccaccac cactgagatc cggctgctaa caaagcccg a
aggaagctg
61  agttggetgc tgccaccgct gagcaataac tagcataacc cttggggcc t
ctaaacggg
121  tcttgagggg ttttttgcg aaaggaggaa ctatatccg attggcgaat g
ggacgcgcc
181  ctgtagcggc gcattaagcg cggcgggtgt ggtggttacg cgcagcgtga c
cgctacact
241  tgccagcgcc ctacgcgcc ctcttttcgc tttcttcct tcctttctcg c
cacgttcgc
301  cggctttccc cgtcaagctc taaatcggg gctccctta gggttccgat t
tagtgcttt
361  acggcacctc gacccaaaa aacttgatta gggtagtgt tcacgtagt g
gccatcgcc
421  ctgatagaag gtttttcgcc ctttgacgtt ggagtccaag tctttaata g
tggactctt
481  gttccaaaact ggaacaacac tcaaccctat ctccgtctat tcttttgatt t
ataagggat
541  tttgcgatt tcggcctatt ggttaaaaa tgagctgatt taacaaaaat t
taacgcgaa
601  ttttaacaaa atattaacgt ttacaatttc aggtggcact tttcggggaa a
tgtgcgcgg
661  aaccctatt tgtttatttt tctaaataca ttcaaatatg tatcgcctca t
gaattaatt
721  cttagaaaaa ctcatcgagc atcaaatgaa actgcaattt attcatatca g
gattatcaa
781  taccatattt ttgaaaaagc cgtttctgta atgaaggaga aaactcaccg a
ggcagttcc
841  ataggatggc aagatcctgg tatoggtctg cgattccgac tcgtccaaca t
caatacaac
901  ctattaattt ccctcgtca aaaataaggt tatcaagtga gaaatcacca t
gagtgcgca
961  ctgaatccgg tgagaatggc aaaagtttat gcatttcttt ccagacttgt t
caacaggcc
1021  agccattacg ctggtcatca aaatcactcg catcaaccaa accgttatto a
ttcgtgatt
1081  gcgcctgagc gagacgaaat acgcgatcgc tgttaaaagg acaattacaa a
caggaatcg
1141  aatgcaaccg gcgcaggaac actgccagcg catcaacaat attttcacct g
aatcaggat
1201  attcttctaa tacctggaat gctgttttcc cggggatcgc agtgggtgag a
accatgcat

```

FIGURE 16 I

WO 02/090300

PCT/US02/14358

```

1261 catcaggagt acggataaaa tgcttgatgg tcggaagagg cataaattcc g
tcagccagt
1321 ttagtctgac catctcatct gtaacatcat tggcaacgct acctttgcca t
gtttcagaa
1381 acaactctgg cgcctcgggc ttcccatata atcgatagat tgctgcacct g
attgcccga
1441 cattatogcg agccatttta taccatata aatcagcatc catgttgga t
ttaatcgcg
1501 gcctagagca agacgtttcc cgttgaatat ggctcataac accccttgta t
tactgttta
1561 tgtaagcaga cagttttatt gttcatgacc aaaatccctt aacgtgagtt t
tcgttcac
1621 tgagcgtcag accccgtaga aaagatcaaa ggatcttctt gagatccttt t
tttctgcgc
1681 gtaatctgct gcttgcaaac aaaaaaacca ccgtaccag cgggtggtttg t
ttgcggat
1741 caagagctac caactctttt tccgaaggta actggcttca gcagagcgca g
ataccaaat
1801 actgtccttc tagttagacc gtagttaggc caccacttca agaactctgt a
gcaccgct
1861 acatacctcg ctctgctaata cctgttacca gtggctgctg ccagtggcga t
aagtcgtgt
1921 cttaccgggt tggaactcaag acgatagtta ccggataagg cgcagcggtc g
ggctgaacg
1981 ggggggtcgt gcacacagcc cagcttgagg cgaacgacct acaccgaact g
agataccta
2041 cagcgtgagc tatgagaaag cgcacgctt cccgaaggga gaaaggcgga c
aggatccg
2101 gtaagcggca gggtcggaac aggagagcgc acgagggagc ttccaggggg a
aacgcctgg
2161 tatctttata gtctgtcgg gtttcgccac ctctgacttg agcgtcgatt t
ttgtgatgc
2221 tcgtcagggg ggcggagcct atggaaaaac gccagcaacg cggccttttt a
cggttcctg
2281 gcccttttgc gcccttttgc tcacatgttc tttcctgctg tatccctga t
tctgtggat
2341 aaccgtatta ccgcctttga gtgagctgat accgctcgcc gcagccgaac g
accgagcgc
2401 agcgagtcag tgagcgagga agcggaagag cgcctgatgc ggtattttct c
cttacgcat
2461 ctgtgcggta ttccacacgc catatatggt gcactctcag tacaatctgc t
ctgatgccg
2521 catagttaag ccagtataca ctccgctatc gctacgtgac tgggtcatgg c
tgcccccgc
2581 acaccgcga acaccgctg acgcgcctg acgggcttgt ctgctcccg c
atccgctta
2641 cagacaagct gtgaccgtct ccgggagctg catgtgtcag aggttttcac c
gtcatcacc

```

FIG 16 I (CONT)

66/113

WO 02/090300

PCT/US02/14358

```

2701 gaaacgcgcg aggcagctgc ggtaaagctc atcagcgtgg tcgtgaagcg a
ttcacagat
2761 gtctgcctgt tcatccgctt ccagctcgtt gagtttctcc agaagcgta a
tgtctggct
2821 tctgataaag cgggccatgt taaggcggtt ttttctctgt ttggtcactg a
tgcctccgt
2881 gtaaggggga tttctgttca tgggggtaat gataccgatg aaacgagaga g
gatgctcac
2941 gatacgggtt actgatgatg aacatgcccg gttactggaa cgttgtagag g
taaacact
3001 ggcggtatgg atcgccggg accagagaaa aatcactcag ggtcaatgcc a
gcgcttctg
3061 taatacagat gtagggtgtt cacagggtag ccagcagcat cctgcgatgc a
gatccgga
3121 cataatgggt cagggcgtg acttccgctt ttccagactt tacgaaacac g
gaaacgaa
3181 gaccattcat gttgttctc aggtcgaga cgttttgtag cagcagtcgc t
tcacgttcg
3241 ctgcgtatc ggtgattcat totgctaacc agtaagcaa ccccgccagc c
tagccgggt
3301 cctcaacgac aggagcacga tcatgcgcac ccgtggggcc gccatgccgg c
gataatggc
3361 ctgcttctcg cgaacagtt tgggtggcgg accagtgacg aaggcttgag c
gagggcgtg
3421 caagattccg aataccgaa cgcacagggc gatcatctgc gcgctccagc g
aaagcgctc
3481 ctcccgaaa atgaccaga gcgctgcgag cacctgtcct acgagttgca t
gataaagaa
3541 gacagtcata agtgccgga cgatagtcac gcccgcgcc caccggaagg a
gctgactgg
3601 gttgaaggct ctcaagggca tcggtcgaga tcccggtgcc taatgagtga g
ctaacttac
3661 attaatgctg ttgcgctcac tgcccgcttt ccagtcggga aacctgtgt g
ccagctgca
3721 ttaatgaatc ggccaacgcg cggggagagg cggtttgctg attgggcgcc a
gggtggttt
3781 ttcttttcac cagtgcagc ggcaacagct gattgccctt caccgcctgg c
cctgagaga
3841 gttgcagcaa gcggtccacg ctggtttgcc ccagcaggcg aaaatcctgt t
tgatgggtg
3901 ttaacggcgg gatataacat gagctgtctt cggtatcgtc gtatcccact a
ccgagatgt
3961 ccgcaccaac gcgcagcccg gactcggtaa tggcacgcat tgcgccagc g
ccatctgat
4021 cgttgccaac cagcatcgca gtgggaacga tgcctcatt cagcatttgc a
tggtttgtt
4081 gaaaaccgga catggcactc cagtgcctt cccgttccgc tatcggtga a
tttgattgc

```

FIG 16 I (CONT)

WO 02/090300

PCT/US02/14358

```

4141 gagtgagata ttatgccag ccagccagac gcagacggcg cgagacagaa c
ttaatgggc
4201 ccgctaacag cgcgatttgc tggtagacca atgcgaccag atgctccacg c
ccagtcgcg
4261 taccgctctt atgggagaaa ataatactgt tgatgggtgt ctggtcagag a
catcaagaa
4321 ataacgccg aacattagt caggcagctt ccacagcaat ggcacccctg t
catccagcg
4381 gatagttaat gatcagccca ctgacgcgtt gcgcgagaag attgtgcacc g
ccgctttac
4441 aggccttcgac gccgcttctg tctaccatcg acaccaccac gctggcacc a
gttgatcgg
4501 cgcgagattt aatcgccgag acaatttgcg acggcgcggt cagggccaga c
tggaggtgg
4561 caacgccaat cagcaacgac tggttgccg ccagttgttg tgccacgcgg t
tgggaatgt
4621 aatcagctc cgcctatcg gcttccactt tttcccgct tttcgcagaa a
cgtggctgg
4681 cctggttcac cagcgggaa acggtctgat aagagacacc ggcatactct g
cgacatcgt
4741 ataacgttac tggtttcaca ttcaccacc tgaattgact ctcttcggg c
gctatcatg
4801 ccataccgag aaagggtttg cgcctatcga tgggtgccg gatctcga c
tctccctta
4861 tgcgactcct gcattaggaa gcagccagc agtaggttga ggcggttgag c
accgcccgc
4921 gcaaggaatg gtgcctgcaa ggagatggcg cccaacagtc ccccgccac g
gggcctgcc
4981 accataccca cgcgaaaca agcgtctatg agcccgaggt ggcgagccc a
tctcccca
5041 tcggtgatgt cggcgatata ggcgccagca accgcacctg tggcgccggt g
atgcgggcc
5101 acgatgcgtc cggcgtagag gatcgagatc gatctcgatc ccgcgaaatt a
atacgactc
5161 actatagggg aattgtgagc ggataacaat tccctctag aaataatttt g
tttaacttt
5221 aagaaggaga tatacatatg aaagaaaccg ctgctgctaa attcgaacgc c
agcacatgg
5281 acagcccaga tctgggtacc ctggtgccac gcggttccat ggctgatatc a
gatctaag
5341 gtctgaaac tcacaacaca aggtctgtga tcttaggaag tggcccagcg g
cacacacgg
5401 cggcgattta cgcagctagg gctgaactta aacctcttct cttcgaagga t
ggatggcta
5461 acgacatcgc tcccggtggt caactaaca ccaaccacga cgtcgagaat t
tcccggat
5521 ttccagaagg tattctcgga gtagagctca ctgacaaatt ccgtaacaa t
cggagcgat

```

FIG 16 I (CONT)

68/113

WO 02/090300

PCT/US02/14358

5581 tgggtactac gatatttaca gagacgggtga cgaaagtoga tttctcttcg a
aaccgttta
5641 agctattcac agattccaag gccattctcg ctgacgctgt gattctcgct a
ctggagctg
5701 tggctaagcg gcttagcttc gttggatctg gtgaagggttc tggagggttc t
ggaaccgtg
5761 gaatctccgc atgcgctggt tgcgacggag ctgctccgat attccgtaac a
aacctcttg
5821 cgggtgatcgg tggaggcgat tcagcaatgg aagaagcaaa ctttcttaca a
aatatggat
5881 ccaaagtgtg tataatccat cgctttgatg cttttaacgc gtctaagatt a
tgcagcagc
5941 gcgctttgtc taatcctaag attgatgtga tttggaactc gtctgttggtg g
aagcttatg
6001 gagatggaga aagagatgtg ctggaggat tgaagtga gaatgtggtt a
ccggtgatg
6061 tttctgattt aaaagtttct ggattgttct ttgctattgg tcatgagcca g
ctaccaagt
6121 ttttgatgg tgggtgtgag ttagattcgg atggttatgt tgtcacgaag c
ctggacta
6181 cacagactag cgttcccga gtttctgctg cgggtgatgt tcaggataag a
agtataggc
6241 aagccatcac tgctgcagga actgggtgca tggcagcttt ggaatgcagag c
attacttac
6301 aagagattgg atctcagcaa ggtaagagtg atggagtoga caagcttgcg g
ccgcac

FIG 16 I (CONT)

WO 02/090300

PCT/US02/14358

The RRR-WT variant coding sequence with S-tag at the N-terminus, His-Tag at C-terminus (5238-26)

```

1  tcgagcacca ccaccaccac cactgagatc cggctgctaa caaagcccga a
aggaagctg
61  agttggctgc tgccaccgct gagcaataac tagcataacc ccttggggcc t
ctaaacggg
121  tcttgagggg ttttttgctg aaaggaggaa ctatatccgg attggcgaat g
ggacgcggc
181  ctgtagcggc gcattaagcg cggcgggtgt ggtggttacg cgcagcgtga c
cgctacact
241  tgccagcgcc ctacgcgccg ctcccttcgc tttcttccct tcctttctcg c
cacgttcgc
301  cggcttttccc cgtcaagctc taaatcgggg gctcccttta gggttccgat t
tagtgcttt
361  acggcacctc gaccccaaaa aacttgatta gggtagtggt tcacgtagtg g
gccatcgcc
421  ctgatagacg gtttttcgcc ctttgacgtt ggagtccacg ttctttaata g
tggactctt
481  gttccaaaact ggaacaacac tcaaccctat ctcggtctat tcttttgatt t
ataagggat
541  ttgccgattt tcggcctatt gggttaaaaa tgagctgatt taacaaaaat t
taacgcgaa
601  ttttaacaaa atattaacgt ttacaatttc aggtggcoact tttcggggaa a
tgtgcgcgg
661  aacccctatt tgtttatttt tctaaataca ttcaaatacg tatccgctca t
gaattaatt
721  cttagaaaaa ctcatcgagc atcaaatgaa actgcaattt attcatatca g
gattatcaa
781  taccatattt ttgaaaaagc cgtttctgta atgaaggaga aaactcaccg a
ggcagttcc
841  ataggatggc aagatcctgg tatcggtctg cgattccgac tcgtccaaca t
caatacaac
901  ctattaattt ccctcgtca aaaataaggt tatcaagtga gaaatcacca t
gagtgcgca
961  ctgaatccgg tgagaatggc aaaagtattt gcatttcttt ccagacttgt t
caacaggcc
1021  agccattacg ctcgtcacga aaatcactcg catcaaccaa accgttatcc a
ttcgtgatt
1081  ggcctgagc gagacgaaat acgcgacgcg tgttaaaagg acaattacaa a
caggaatcg
1141  aatgcaaccg ggcagggaac actgccagcg catcaacaat attttcacct g
aatcaggat
1201  attcttctaa tacctggaat gctgttttcc cggggatcgc agtgggtgagt a
accatgcat
1261  catcaggagt acggataaaa tgcttgatgg tcggaagagg cataaattcc g

```

FIGURE 16J

70/113

WO 02/090300

PCT/US02/14358

tcagccagt
1321 ttagtctgac catctcatct gtaacatcat tggcaacgct acctttgcc a t
gtttcagaa
1381 acaactctgg cgcctcgggc ttcccatata atcgatagat tgtcgcacct g
attgccga
1441 cattatcgcg agccattta taccatata aatcagcatc catgttgaa t
ttaatgcg
1501 gcctagagca agacgtttcc cgttgaatat ggctcataac accccttgta t
tactgttta
1561 tgaagcaga cagttttatt gttcatgacc aaaatccctt aacgtgagtt t
tcgttcac
1621 tgagcgtcag acccctaga aaagatcaaa ggatcttctt gagatccttt t
ttctcgcg
1681 gtaatctgct gcttgcaaac aaaaaaaca ccgctaccag cgggtggttg t
ttgcggat
1741 caagagctac caactctttt tcogaaggta actggcttca gcagagcgca g
ataccaaat
1801 actgtccttc tagttagcc gtagttaggc caccacttca agaactctgt a
gcaccgct
1861 acatacctog ctctgctaact cctgttacca gtggctgctg ccagtggcga t
aagtcgtgt
1921 cttaaccgggt tggactcaag acgatagta ccggataagg cgcagcggtc g
ggctgaacg
1981 gggggttctg gcacacagcc cagcttgagc cgaacgacct acaccgaact g
agataccta
2041 cagcgtgagc tatgagaaag cgcacagctt ccgaaggga gaaaggcgga c
aggatccg
2101 gtaagcggca gggcgggaac aggagagcgc acgagggagc ttccaggggg a
aacgcctgg
2161 tatctttata gtctgtcgg gtttcgccac ctctgacttg agcgtcgatt t
ttgtgatgc
2221 tcgtcagggg ggcggagcct atggaaaaac gccagcaacg cggccttttt a
cggctcctg
2281 gcccttttgc ggccttttgc tcacatgttc tttcctgctg tatccctga t
tctgtggat
2341 aaccgtatta ccgcctttga gtgagctgat accgctcgcc gcagccgaac g
accgagcgc
2401 agcaggtcag tgagcgagga agcgggaagc cgcctgatgc ggtattttct c
cttaacgat
2461 ctgtgoggta ttacacacc catatatggt gcaactctcag tacaatctgc t
ctgatgccg
2521 catagttaag ccagtataca ctccgctatc gctacgtgac tgggtcatgg c
tgcgcccg
2581 acaccgccca acaccgctg acgcgccctg acgggcttgt ctgctccgg c
atccgctta
2641 cagacaagct gtgacgtct ccgggagctg catgtgtcag aggttttcac c
gtcatcacc
2701 gaaacgcgcg aggcagctgc ggtaaaagctc atcagcgtgg tcgtgaagcg a

FIG 16J (CONT)

WO 02/090300

PCT/US02/14358

ttcacagat
2761 gtctgcctgt tcaccgcgt ccagctcgtt gagtttctcc agaagcgta a
tgtctggct
2821 tctgataaag cgggccatgt taaggcggt ttttctctgt ttggctactg a
tgctccgt
2881 gtaagggga tttctgttca tgggggtaat gataccgatg aaacgagaga g
gatgctcac
2941 gatacgggtt actgatgatg aacatgccg gttactgaa cgttgtagg g
taaacact
3001 ggcggtatgg atgcggcgg accagagaaa aatcactcag ggtcaatgcc a
gcgcttcgt
3061 taatacagat gtaggtgttc cacagggtag ccagcagcat cctgcgatgc a
gatccgaa
3121 cataatggtg caggcgctg acttcgcgt ttccagactt tacgaaacac g
gaaacgaa
3181 gaccattcat gttgtgttc aggtgcaga cgttttgtag cagcagtcg t
tcacgttcg
3241 ctccgctatc ggtgatcat tctgtaacc agtaaggcaa ccccgccagc c
tagcgggt
3301 cctcaacgac agggagcaga tcatgcgcac ccgtggggcc gccatgccg c
gataatggc
3361 ctgcttctcg ccgaaacgtt tgggtggcgg accagtgcg aagccttgag c
gaggcggtg
3421 caagattccg aataccgcaa gcgacaggcc gatcatcgtc gcgctccagc g
aaagcggtc
3481 ctcccgaaa atgaccaga gcgctgcgg cactgtcct acgagttgca t
gataaagaa
3541 gacagtcata agtgcggcga cgtatgcat gcccgccgcc caccggaagg a
gctgactgg
3601 gttgaaggct ctcaagggca tcggtcgaga tcccggtgcc taatgagtga g
ctaacttac
3661 attaattgcg ttgcgctcac tgcccgcttt ccagtcggga aacotgtcgt g
ccagctgca
3721 ttaatgaatc ggccaacgag cggggagagg cggtttgctg attgggcgcc a
gggtggttt
3781 ttcttttcac cagtgcgacg ggcaacagct gattgcctt caccgcttg c
cctgagaga
3841 gttgcagcaa gcggtccacg ctggtttgcc ccagcaggcg aaaatcctgt t
tgatggtgg
3901 ttaacggcgg gatataacat gagctgtctt cggtatcgtc gtatccact a
ccgagatgt
3961 ccgcaccaac gcgcagcccg gactcggtaa tggcacgcat tgcgcccagc g
ccatctgat
4021 cgttggaac cagcatcgca gtgggaacga tgccctcatt cagcatttgc a
tggtttgtt
4081 gaaaaccgga catggcactc cagtcgcctt ccggttccgc tatcggctga a
tttgattgc
4141 gagtgcagata tttatgccag ccagccagac gcagacgcgc cgagacagaa c

FIG 16J (CONT)

WO 02/090300

PCT/US02/14358

```

ttaatgggc
4201 ccgctaacag cgcgatttgc tggtagacca atgcgaccag atgctccacg c
ccagtcggg
4261 taccgctcttc atgggagaaa ataatactgt tgatgggtgt ctggtcagag a
catcaagaa
4321 ataacgccgg aacattagtgc caggcagctt ccacagcaat ggcatcctgg t
catccagcg
4381 gatagttaat gacagacca ctgacgcgtt gcgcgagaag attgtgcacc g
ccgctttac
4441 aggotctgac gcgccttcgt tctaccatcg acaccaaccac gctggcacc a
gttgatcgg
4501 cgcgagattt aatcgccggc acaatttgcg accgcgcgtg cagggccaga c
tggaggtgg
4561 caacgccaat cagcaacgac tgtttgccg ccagttgttg tgccacggg t
tgggaatgt
4621 aattcagctc cgcacatgcc gcttccactt ttcccgcggt ttccgcagaa a
cgtggctgg
4681 cctggttcac cagcgggaa acggtctgat aagagacacc ggcatactct g
cgacatcgt
4741 ataacgttac tggtttcaca ttcaccaccc tgaattgact ctotccggg c
gctatcatg
4801 ccatacccg cgaagttttg cgcacatcga tgggtgcgg gatctcgacg c
tctccctta
4861 tgagactcct gcattaggaa gcagccacgt agtaggttga ggccgttgag c
accgcccgc
4921 gcaaggaatg gtgcatgcaa ggagatggcg cccaacagtc ccccgccac g
ggccctgcc
4981 accataccca cgcgaaaca agcgctcatg agcccgaaat ggccgagccg a
tctcccca
5041 tcggtgatgt cgcgcatata ggcccgaca accgcacctg tggcgccggt g
atgcgggcc
5101 acgatgcgtc cgcgtagag gatcgagatc gatctcgatc cgcgaaatt a
atacgaactc
5161 actatagggg aattgtgagc ggataacaat tccctctag aaataatttt g
tttaacttt
5221 aagaaggaga tatacatatg aaagaaccg ctgctgctaa attcgaacgc c
agcacatgg
5281 acagcccaga tctgggtacc ctggtgccac gcggttccat ggcgatata a
gatctaag
5341 gtctcgaac tcacaacaca aggcctctgta tcgtaggag tggccagcg g
cacacacgg
5401 cggcgattta cgcagctagg gctgaactta aacctctct ctccgaagga t
ggatggcta
5461 acgacatgc tcccggtggt caactaaca ccaccacga cgtcgagaat t
tcccggtat
5521 ttccagaag tattctcgga gtagagctca ctgacaaatt ccgtaacaa t
cggagcgat
5581 tcggtactac gatatttaca gagacggtga cgaagtcga ttctctctog a

```

FIG 16J (CONT)

WO 02/090300

PCT/US02/14358

aaccgttta
5641 agctattcac agattccaag gccattctcg ctgacgctgt gattctcgct a
ctggagctg
5701 tggctaagcg gcttagcttc gttgcatctg gtgaaggttc tggagggttc t
ggaaccgtg
5761 gaatctccgc atgcgctggt tgcgacggag ctgctccgat attccgtaac a
aacctcttg
5821 cggatgatcg tggaggcgat tcagcaatgg aagaagcaaa ctttcttaca a
aatatggat
5881 ccaaagtgtg tataatccat cgcgcgatg ctttctgtgc gtctaagatt a
tgcagcagc
5941 gcgctttgtc taatcctaag attgatgtga tttggaactc gtctgttgtg g
aagcttatg
6001 gagatggaga aagagatgtg cttggaggat tgaaagtga gaatgtggtt a
cgggtgatg
6061 tttctgattt aaaagtttct ggattgttct ttgctattgg tcatgagcca g
ctaccaagt
6121 ttttggatgg tgggtttgag ttagattcgg atggttatgt tgtcacgaag c
ctggtacta
6181 cacagactag cgttcccga gtttctgctg cgggtgatgt tcaggataag a
agtataggc
6241 aagccatcac tgctgcagga actgggtgca tggcagcttt ggatgcagag c
attacttac
6301 aagagattgg atctcagcaa ggtaagagt atggagtcca caagcttgcg g
ccgcac

FIG 16J (CONT)

74/113

WO 02/090300

PCT/US02/14358

The WVG variant coding sequence with S-tag at the N-terminus (5238-6335)

```

1  tcgagcacca ccaccaccac cactgagatc cggctgctaa caaagcccca a
aggaagctg
61 agttggctgc tgccaccgct gagcaataac tagcataacc ccttggggcc t
ctaaacggg
121 tcttgagggg ttttttgctg aaaggaggaa ctatatccgg attggcgaat g
ggacgcgcc
181 ctgtagcggc gcattaagcg cggcgggtgt ggtggttacg cgcagcgtga c
cgctacact
241 tgccagcgcc ctagcgcccg ctcctttcgc tttcttcctt tcctttctcg c
cacgttcgc
301 cggctttccc cgtcaagctc taaatcgggg gctcccttta gggttccgat t
tagtgcttt
361 acggcacctc gaccccaaaa aacttgatta gggtagtggt tcacgtagtg g
gccatcgcc
421 ctgatagacg gtttttcgcc cttgacgtt ggagtccaog ttccttaata g
tggactctt
481 gttccaaaact ggaacaacac tcaacctat ctcggtctat tcttttgatt t
ataagggat
541 tttgcgatt tcggcctatt ggttaaaaaa tgagctgatt taacaaaaat t
taacgcgaa
601 ttttaacaaa atattaacgt ttacaatttc aggtggcact tttcggggaa a
tgtgcgcgg
661 aacccctatt tgtttathtt tctaaataca ttcaaatacg tatccgctca t
gaattaatt
721 cttagaaaaa ctcacgcagc atcaaatgaa actgcaattt attcatatca g
gattatcaa
781 taccatattt ttgaaaaagc cgtttctgta atgaaggaga aaactcaccg a
ggcagttcc
841 ataggatggc aagatcctgg tatcggctcg cgattccgac tcgtccaaca t
caatacaac
901 ctattaattt cccctcgtca aaaataaggt tatcaagtga gaaatcacca t
gagtgcgca
961 ctgaatccgg tgagaatggc aaaagtttat gcatttcttt ccagacttgt t
caacaggcc
1021 agccattacg ctcgatcatca aaatcactcg catcaaccaa accgttatcc a
ttcgtgatt
1081 gcgcctgagc gagacgaaat acgcgatcgc tgttaaaagg acaattacaa a
caggaatcg
1141 aatgcaaccg gcgcaggaac actgccagcg catcaacaat attttcacct g
aatcaggat
1201 attcttctaa tacctggaat gctgttttcc cggggatcgc agtggtgagt a
accatgcat
1261 catcaggagt acggataaaa tgcttgatgg tcggaagagg cataaattcc g

```

FIGURE 16K

WO 02/090300

PCT/US02/14358

tcagccagt
 1321 ttagtctgac catctcatct gtaacatcat tggcaacgct acotttgcca t
 gtttcagaa
 1381 acaactctgg cgcctcgggc ttcccatata atcgatagat tgctgcacct g
 attgcccga
 1441 cattatcgcg agcccattta taccatata aatcagcacc catgttgga t
 ttaatcgcg
 1501 gcctagagca agacgtttcc cgttgaatat ggctcataac accccttgta t
 tactgttta
 1561 tgtaagcaga cagttttatt gttcatgacc aaaatccott aacgtgagtt t
 tcgttccac
 1621 tgagcgtcag acccgttaga aaagatcaaa ggatcttctt gagatccttt t
 tttctgcgc
 1681 gtaatctgct gcttgcaaac aaaaaacca ccgctaccag cgggtggtttg t
 ttgcccgat
 1741 caagagctac caactctttt tccgaaggta actggcttca gcagagcgca g
 ataccaaat
 1801 actgtccttc tagtgtagcc gtagttagcc caccacttca agaactctgt a
 gcaccgct
 1861 acatacctcg ctctgctaata cctgttacca gtggctgctg ccagtgcgca t
 aagtctgt
 1921 cttaccgggt tggactcaag acgatagtta ccggataagg cgcagcggtc g
 ggctgaacg
 1981 gggggttcgt gcacacagcc cagcttgag cgaacgaact acaccgaact g
 agataccta
 2041 cagcgtgagc tatgagaaag cgcacgcctt ccgaaggga gaaaggcgga c
 aggtatccg
 2101 gtaagcggca gggctggaac aggagagcgc acgaggagc ttccaggggg a
 aacgcctgg
 2161 tatctttata gtctgtcgg gtttcgcac ctctgacttg agcgtcgatt t
 ttgtgatgc
 2221 tcgtcagggg ggogagcct atggaaaac gccagcaacg cggccttttt a
 cggttcctg
 2281 gccttttctt ggcttttgc tcacatgttc tttctgcgt tatccctga t
 tctgtggat
 2341 aacggtatta ccgccttga gtgagctgat accgctcgc ccagccgaac g
 accgagcgc
 2401 agcgagtcag tgagcgagga agcggaagag cgcctgatgc ggtattttct c
 cttacgcat
 2461 ctgtgoggta tttcacaccg catatatggt gcaactcag tacaatctgc t
 ctgatgcg
 2521 catagttaag ccagtataca ctccgctatc gctacgtgac tgggtcatgg c
 tgcgcccgc
 2581 acaccgccca acaccgctg acgcgccctg acgggcttgt ctgctcccg c
 atccgctta
 2641 cagacaagct gtgaccgtct ccggagctg catgtgtcag aggttttcc c
 gtcacacc
 2701 gaaacgcgcg aggcagctgc ggtaaagctc atcagcgtgg tcgtgaagcg a

FIG 16K (CONT)

WO 02/090300

PCT/US02/14358

ttcacagat
2761 gtctgcctgt tcatccgcgt ccagctcgtt gagttctcc agaagcgtta a
tgtctggct
2821 tctgataaag cgggccatgt taaggcggt ttttctctgt ttggtcactg a
tgcctccgt
2881 gtaagggga tttctgttca tgggggtaat gataccgatg aaacgagaga g
gatgctcac
2941 gatacgggtt actgatgatg aacatgcccg gttactggaa cgttgtgagg g
taaacaact
3001 ggcggtatgg atgcggcggg accagagaaa aatcactcag ggtcaatgcc a
gcgcttcgt
3061 taatacagat gtaggtgttc cacagggtag ccagcagcat cctgcgatgc a
gatccgaa
3121 cataatggg caggcgctg acttccgcgt ttccagactt tacgaaacac g
gaaaccgaa
3181 gaaccattcat gttgttgctc aggtgcaga cgttttgag cagcagtcgc t
tcaogttcg
3241 ctgcggtatc ggtgattcat tctgctaacc agtaaggcaa ccccgccagc c
tagccgggt
3301 cctcaacgac aggagcacga tcatgcgcac ccgtggggcc gccatgccgg c
gataatggc
3361 ctgcttctcg ccgaaacgtt tgggtggcgg accagtgacg aaggcttgag c
gaggcgctg
3421 caagattccg aataccgcaa gcgacaggcc gatcatcgtc gcgctccagc g
aaagcggtc
3481 ctgcgcgaaa atgaccaga gcgctgcgg cactgtctc acgagttgca t
gataaagaa
3541 gacagtcata agtgcggcga cgatagtcac gcccgcgcc caccggaagg a
gctgactgg
3601 gttgaaggct ctcaaggcga tccgtcgaga tcccggtgcc taatgagtga g
ctaacttac
3661 attaatggcg ttgcgctcac tgcccgcttt ccagtcggga aacctgtcgt g
ccagctgca
3721 ttaatgaatc ggccaacgcg cggggagagg cggtttgctg attgggcgcc a
gggtgggtt
3781 ttcttttcac cagtgcgacg ggcaacagct gattgcccct caccgcctgg c
cctgagaga
3841 gttgcagcaa gcggtccacg ctggtttgcc ccagcaggcg aaaatcctgt t
tgatgggtg
3901 ttaacggcgg gatataacat gagctgtctt cggtatcgtc gtatcccact a
ccgagatgt
3961 ccgaccaaac gcgcagccc gactcggtaa tggcagcat tgcgccagc g
ccatctgat
4021 cgttggcaac cagcatcgca gtgggaacga tgccctcatt cagcatttgc a
tggtttgtt
4081 gaaaaccgga catggcactc cagtgcctt cccgttcgc tateggctga a
tttgattgc
4141 gagtgcagata ttatgccag ccagccagac gcagacgcgc cgagacagaa c

FIG 16K (CONT)

77/113

WO 02/090300

PCT/US02/14358

ttaatgggc
 4201 ccgctaacag cgcgatttgc tggtagacca atgcgaccag atgctccacg c
 ccagtgcgcg
 4261 taccgtcttc atgggagaaa ataatactgt tgaagggtgt ctggtcagag a
 catcaagaa
 4321 ataacgcgcg aacattagtg caggcagctt ccacagcaat ggcatacctg t
 catccagcg
 4381 gatagttaat gatcagccca ctgacgcgtt gcgcgagaag attgtgcacc g
 cgcgtttac
 4441 aggccttcgac gccgcttcgt tctaccatcg acaccaccac gctggcacc a
 gttgatcgg
 4501 cgcgagatatt aatgcgcgcg acaatttgcg acggcgcgtg cagggccaga c
 tggaggtgg
 4561 caacgccaat cagcaacgac tgtttgcccg ccagtgttg tggcacgcg t
 tgggaatgt
 4621 aattcagctc cgcctatcgc gcttccactt tttccgcgt tttcgcagaa a
 cgtggctgg
 4681 cctggttcac cagcgcggaa acggtctgat aagagacacc ggcatactct g
 cgacatcgt
 4741 ataacgttac tggtttcaca ttcaccacc tgaattgact ctctccggg c
 gctatcatg
 4801 ccataccgcg aaaggttttg cgcctatcga tgggttcgag gatctgcag c
 tctccotta
 4861 tgcgactcct gcattaggaa gcagccagc agtaggttga ggcggttgag c
 accgcgcgc
 4921 gcaaggaatg gtgcagcaa ggagatggcg cccaacagtc ccccgccac g
 gggcctgcc
 4981 accataccca cgcgaaaaca agcgtctatg agccgaagt ggcgagcccg a
 tcttccca
 5041 tcggtgatgt cggcgatata ggcgccagca accgcactg tggcgccggt g
 atgcgcggc
 5101 acgatgcgtc cggcgtagag gatcgagatc gatctcgatc cgcggaatt a
 atacgactc
 5161 actatagggg aattgtgagc ggataacaat tcccctctag aaataatttt g
 tttaacttt
 5221 aagaaggaga tatacatatg aaagaaaccg ctgctgctaa attcgaacgc c
 agcacatgg
 5281 acagcccaga tctgggtacc ctggtgccac gcggttccat ggctgatatc a
 gatctaattg
 5341 gtctcgaaac tcacaacaca aggctctgta tcgtaggaag tggccagcg g
 cacacacgg
 5401 cggcgattta cgcagctagg gctgaactta aacctcttct cttcgaagga t
 ggatggcta
 5461 acgacatcgc tcccggtggt caactaaca ccaccaccga cgtcgagaat t
 tcccggat
 5521 ttccagaagg tattctcgga gtagagctca ctgacaaatt ccgtaaaca t
 cggagcgat
 5581 tcggtactac gatatttaca gagacggtga cgaaagtga tttctcttcg a

FIG 16K (CONT)

WO 02/090300

PCT/US02/14358

```
aaccgttta
5641 agctattcac agattccaag gccattctcg ctgacgctgt gattctcgct a
ctggagctg
5701 tggctaagcg gotttagcttc gttggatctg gtgaagggtc tggagggttc t
ggaaccgtg
5761 gaatctccgc atgcgctggt tgcgaaggag ctgctccgat attocgtaac a
aacctcttg
5821 cggatgatcg tggaggcgat tcagcaatgg aagaagcaaa ctttcttaca a
aatatggat
5881 ccaaagtgtg tataatccat tgggtggatg cttttggggc gtctaagatt a
tgcagcagc
5941 gcgctttgtc taatcctaag attgatgtga ttggaactc gtcgttgtg g
aagcttatg
6001 gagatggaga aagagatgtg cttggaggat tgaagtga gaatgtggtt a
ccggtgatg
6061 tttctgattt aaaagtttct ggattgttct ttgctattgg tcatgagcca g
ctaccaagt
6121 ttttggatgg tgggtgttag ttagattcgg atggttatgt tgcacgaag c
ctggtacta
6181 cacagactag cgttcccgga gttttcgctg cgggtgatgt tcaggataag a
agtataggc
6241 aagccatcac tgctgcagga actgggtgca tggcagcttt ggatgcagag c
attacttac
6301 aagagattgg atctcagcaa ggtaagagtg attgagtcga caagcttgcg g
ccgcac
```

FIG 16K (CONT)

WO 02/090300

PCT/US02/14358

The WRS variant coding sequence with S-tag at the N-terminus (5238-6335)

```

1  tcgagcacca ccaccaccac cactgagatc cggctgctaa caaagcccca a
aggaagctg
61 agttggtgtg tgccaccgct gagcaataac tagcataacc ccttggggcc t
ctaaacggg
121 tcttgagggg tttttgtgtg aaaggaggaa ctatatccgg attggcgaat g
ggacgggcc
181 ctgtagcggc gcattaagcg cggcgggtgt ggtgggtacg cgcagcgtga c
cgctacact
241 tgcacggccc ctagcggccc ctcctttcgc ttcttccct tcctttctcg c
cacgttcgc
301 cggctttccc cgtcaagctc taaatcgggg gctccctta gggttccgat t
tagtgcttt
361 acggcacctc gaccccaaaa aaactgatta gggtagtggt tcacgtagtg g
gccatcgcc
421 ctgatatagc gtttttcgcc ctttgacgtt ggagtccacg ttctttaata g
tggactctt
481 gttccaaact ggaacaacac tcaaccctat ctgggtctat tcttttgatt t
ataagggat
541 tttgccgatt tcggcctatt ggttaaaaaa tgagctgatt taacaaaaat t
taacgcgaa
601 ttttaacaaa atattaacgt ttacaatttc aggtggcact tttcggggaa a
tgtgcgcgg
661 aaccctctat tgtttatttt tctaaataca ttcaaatatg tatccgctca t
gaattaatt
721 cttagaaaaa ctcatcgagc atcaaatgaa actgcaattt attcatatca g
gattatcaa
781 taccatatth ttgaaaaagc cgtttctgta atgaaggaga aaactcaccg a
ggcagttcc
841 ataggatggc aagatcctgg tatcgggtcg cgattccgac tcgtccaaca t
caatacaac
901 ctattaattt ccctcgtca aaaaataggt tatcaagtga gaaatcacca t
gagtgcga
961 ctgaatccgg tgagaatggc aaaagtttat gcatttcttt ccagacttgt t
caacaggcc
1021 agccattacg ctgctcatca aaatcactcg catcaaccaa accgttatth a
ttcgtgatt
1081 gcgcctgagc gagacgaaat acgcgatcgc tggtaaaagg acaattacaa a
cagggaatcg
1141 aatgcaaccg gcgcaggaac actgccagcg catcaacaat attttcacct g
aatcaggat
1201 attcttctaa taactggaat gctgttttcc cggggatcgc agtgggtgagt a
accatgcat

```

FIGURE 16L

WO 02/090300

PCT/US02/14358

```

1261 catcaggagt acggataaaa tgcttgatgg tcggaagagg cataaattcc g
tcagccagt
1321 ttagtctgac catctcatct gtaacatcat tggcaacgct acctttgcca t
gtttcagaa
1381 acaactctgg cgcctcgggc ttcccatata atcgatagat tgcgcacct g
attgcccga
1441 cattatecgc agcccattta taccatata aatcagcatc catgttggaa t
ttaatcgcg
1501 gcctagagca agacgtttcc cgttgaatat ggctcataac accccttgta t
tactgttta
1561 tgtaagcaga cagttttatt gttcatgacc aaaatccctt aacgtgagtt t
tcgtccac
1621 tgagcgtcag accccgtaga aaagatcaaa ggatcttctt gagatccttt t
tttctgcgc
1681 gtaatctgct gcttgcaaac aaaaaacca ccgctaccag cggtggtttg t
ttgcggat
1741 caagagctac caactctttt tccgaaggta actggcttca gcagagcgca g
ataccaaat
1801 actgtccttc tagttagacc gtagttaggc caccacttca agaactctgt a
gcaccgct
1861 acatacctcg ctctgctaatt cctgttacca gtggctgctg ccagtggcga t
aagtcgtgt
1921 cttaccgggt tggactcaag acgatagtta cgggataagg cgcagcggtc g
ggctgaacg
1981 gggggttcgt gcacacagcc cagcttgag cgaacgacct acaccgaact g
agataccta
2041 cagcgtgagc tatgagaaag gccacgctt cccgaaggga gaaagcgga c
aggtatccg
2101 gtaagcgga gggtcggaac aggagagcgc acgagggagc ttccaggggg a
aacgcctgg
2161 tatctttata gtctgtcgg gtttcgccac ctctgacttg agcgtcgatt t
ttgtgatgc
2221 tcgtcagggg ggcggagcct atggaaaaac gccagcaacg cggccttttt a
cggttcctg
2281 gccttttgct ggccttttgc tcacatgttc tttctgcgt tatccctga t
tctgtggat
2341 aaccgtatta ccgcctttga gtgagctgat accgctcgcc gcagccgaac g
accgagcgc
2401 agcgagtcag tgagcgagga agcgggaagag cgcctgatgc ggtattttct c
cttaacgat
2461 ctgtgcggta ttccacaccg catatatggt gcactctcag tacaatctgc t
ctgatgccg
2521 catagttaag ccagtataca ctccgtatc gctacgtgac tgggtcatgg c
tgcccccg
2581 acacccgcca acacccgctg acgcgccctg acgggcttgt ctgctcccgg c
atccgctta
2641 cagacaagct gtgaccgtct ccgggagctg catgtgtcag aggttttcac c
gtcatcacc

```

FIG 16L (CONT)

WO 02/090300

PCT/US02/14358

```

2701 gaaacgcgcg aggcagctgc ggtaaagctc atcagcgtgg tcgtgaagcg a
ttcacagat
2761 gtctgcctgt tcateccgct ccagctcgtt gagtttctcc agaagcgta a
tgtctggct
2821 tctgataaag cgggccatgt taaggcgggt tttttcctgt ttggtaactg a
tgcctccgt
2881 gtaaggggga tttctgttca tgggggtaat gataccgatg aaacgagaga g
gatgctcac
2941 gatacgggtt actgatgatg aacatgcccg gttactggaa cgttgtgagg g
taaacact
3001 ggcggtatgg atgcggcggg accagagaaa aatcactcag ggtcaatgcc a
gcgcttctg
3061 taatacagat gtaggtgttc cacagggtag ccagcagcat cctgcgatgc a
gatccgaa
3121 cataatgggt cagggcgctg acttcgcgtt ttccagactt tacgaaacac g
gaaacgaa
3181 gaccattcat gttgttgcct aggtcgcaga cgttttgca cagcagtcgc t
tcacgttcg
3241 ctgcgctatc ggtgattcat tctgctaacc agtaaggcaa ccccgccagc c
tagccgggt
3301 cctcaacgac agggagcacg tcattgcgac ccgtgggggc gccatgcocg c
gataatggc
3361 ctgcttctcg ccgaaacgtt tgggtggcgg accagtgacg aaggcttgag c
gaggcgctg
3421 caagattccg aataccgcaa gcgacaggcc gatcatcgtc gcgctccagc g
aaagcgctc
3481 ctgcgccgaa atgaccaga gcgtgcggc cactgtcct acgagttgca t
gataaagaa
3541 gacagtcata agtgccgca cagatgcat gcccgcgcc caccggaagg a
gctgactgg
3601 gttgaaggct ctcaagggca toggtcgaga tcccggtgcc taatgagtga g
ctaaactac
3661 attaatgctg ttgcgctcac tgcccgcttt ccagtcggga aacctgtcgt g
ccagctgca
3721 ttaatgaatc ggccaaocg cggggagagg cggtttgct attggcgcc a
gggtggttt
3781 ttcttttcc cagtgaagc ggcaacagct gattgccctt caccgcctgg c
cctgagaga
3841 gttgcagcaa gcggtccacg ctggtttgoc ccagcaggcg aaaatcctgt t
tgatggtgg
3901 ttaacggcgg gatataacat gagctgtctt cggtatogtc gtatccact a
ccgagatgt
3961 ccgcaccaac gcgcagccc gactcggtaa tggcacgcat tgcgccagc g
ccatctgat
4021 cgttggcaac cagcatcgca gtgggaacga tgcctcatt cagcatttgc a
tggtttgtt
4081 gaaaaccgga catggcactc cagtcgcctt cccgttcgc tatoggtga a
tttgattgc

```

FIG 16L (CONT)

WO 02/090300

PCT/US02/14358

```

4141 gagtgagata ttatgccag ccagccagac gcagacgcgc cgagacagaa c
ttaatgggc
4201 ccgctaacag cgcgatttgc tggtagacca atgcgaccag atgtccacg c
ccagtcgcg
4261 tacogtcttc atgggagaaa ataatactgt tgatgggtgt ctggtcagag a
catcaagaa
4321 ataacgccg aacattagt caggcagctt ccacagcaat ggcacctgg t
catccagcg
4381 gatagttaat gatcagcca ctgacgcgtt gcgcgagaag attgtgcacc g
ccgctttac
4441 aggccttcgac gcgcttcgt tctaccatcg acaccaccac gctggcacc c
gttgatcgg
4501 cgcgagattt aatgcgcgcg acaatttgcg acggcgcggt cagggccaga c
tggagggtg
4561 caacgccaat cagcaacgac tgtttgccg ccagttgttg tggcacgcg t
tgggaatgt
4621 aattcagctc cgcacatgcc gcttcacatt ttcccgctt ttccgagaa a
cgtggctgg
4681 cctgggtcac cagcgggaa acggtctgat aagagacacc ggcatactct g
cgacatcgt
4741 ataacgttac tggtttcaca ttcaccacc tgaattgact ctctccggg c
gctatcatg
4801 ccataccgcg aaagggtttg cgcacatcga tgggtgtccg gatctcgacg c
tctccctta
4861 tgcgactctc gcattaggaa gcagcccagt agtaggttga ggccgttgag c
accgcgccc
4921 gcaaggaatg gtgcatgcaa ggagatggcg cccaacagtc ccccgccac g
gggcctgcc
4981 accataccca cgcgaaaca agcgtctcatg agcccgaaat ggcgagccc a
tctcccca
5041 tcggtgatgt cggcgatata ggcccgagca accgcacctg tggcgccggt g
atgcgggcc
5101 acgatgcgtc cggcgtagag gatcgagatc gatctcgatc ccgcgaaatt a
atagactc
5161 actatagggg aattgtgagc ggataacaat tccctctag aaataatttt g
tttaacttt
5221 aagaaggaga tatacatatg aaagaaaccg ctgctgctaa attcgaacgc c
agcacatgg
5281 acagcccaga tctgggtacc ctggtgccac gcggttccat ggctgatatc a
gatctaag
5341 gtctcgaaac tcacaacaca aggtctgtga tcgtaggaag tggcccagcg g
cacacacgg
5401 cggcgattta cgcagctagg gctgaactta aacctctct ctccgaagga t
ggatggcta
5461 acgacatcgc tcccggtggg caactaaca ccaccaccga cgtcgagaat t
tcccggat
5521 ttccagaagg tattctcgga gtagagctca ctgacaaatt ccgtaacaa t
cggagcgat

```

FIG 16ⁱ (CONT)

83/113

WO 02/090300

PCT/US02/14358

5581 tcggtactac gatatttaca gagacggtga cgaaagtcga tttctcttcg a
aaccgttta
5641 agctattcac agattccaag gccattctcg ctgacgctgt gattctcgct a
ctggagctg
5701 tggctaagcg gcttagcttc gttggatctg gtgaaggttc tggaggtttc t
ggaaccgtg
5761 gaatctccgc atgcgctgtt tgcgacggag ctgctccgat attccgtaac a
aacctcttg
5821 csgtgatcgg tggaggcgat tcagcaatgg aagaagcaaa ctttcttaca a
aatatggat
5881 ccaaagtga tataatccat tggagggatg cttttagtgc gtctaagatt a
tgcagcagc
5941 gcgctttgtc taatcctaag attgatgtga ttgggaactc gtctgttggt g
aagcttatg
6001 gagatggaga aagagatgtg cttggaggat tgaaagtga gaatgtggtt a
cgggtgatg
6061 tttctgattt aaaagtttct ggattgttct ttgctattgg tcatgagcca g
ctaccaagt
6121 ttttggatgg tgggtgtgag ttagattcgg atggttatgt tgtcacgaag c
ctggtacta
6181 cacagactag cgttcccgga gttttcgtg cgggtgatgt tcaggataag a
agtataggc
6241 aagccatcac tgctgcagga actgggtgca tggcagcttt ggatgcagag c
attacttac
6301 aagagattgg atctcagcaa ggtaagagtg attgagtcga caagcttgct g
ccacac

FIG 16L (CONT)

84/113

WO 02/090300

PCT/US02/14358

The WFQ variant coding sequence with S-tag at the N-terminus (5238 ; -6335)

```

1  tcgagcacca ccaccaccac cactgagatc cggctgctaa caaagccoga a
aggaagctg
61  agttggctgc tgccaccgct gagcaataac tagcataacc ccttggggcc t
ctaaacggg
121  tcttgagggg ttttttctg aaaggaggaa ctatatccgg attggcgaat g
ggacgcgcc
181  ctgtagcggc gcattaagcg cggcgggtgt ggtgggttaog cgcagcgtga c
cgctacact
241  tgccagcgcc ctacgcgcc ctcccttcgc tttcttccot tcctttctcg c
cacgttcgc
301  cggctttccc cgtcaagctc taaatcgggg gctcccttta gggttccgat t
tagtgcttt
361  acggcacctc gaccccaaaa aacttgatta gggatgaggt tcacgtagtg g
gccatcgcc
421  ctgatagacg gtttttcgcc ctttgacgtt ggagtccagc ttctttaata g
tggactcct
481  gttccaaact ggaacaacac tcaaccctat ctcggtctat tcttttgatt t
ataagggat
541  tttgccgatt tcggcctatt ggtaaaaaa tgagctgatt taacaaaaat t
taacgcgaa
601  ttttaacaaa atattaacgt ttacaatttc aggtggcact tttcggggaa a
tgtgcgcgg
661  aacccctatt tgtttatttt tctaaatata ttcaaatatg tatccgctca t
gaattaatt
721  cttagaaaaa ctcatcgagc atcaaatgaa actgcaattt attcatatca g
gattatcaa
781  taccatattt ttgaaaaagc cgtttctgta atgaaggaga aaactcaccg a
ggcagttcc
841  ataggatggc aagatcctgg taccggtctg cgattccgac tcgtccaaca t
caatacaac
901  ctattaattt ccctctgtca aaaataaggt tatcaagtga gaaatcacca t
gagtgcgca
961  ctgaatccgg tgagaatggc aaaagtttat gcatttcttt ccagacttgt t
caacaggcc
1021  agccattacg ctcgatcata aaatcactcg catcaaccaa accgttattc a
ttcgtgatt
1081  gcgcctgagc gagacgaaat acgcgategc tgtaaaagg acaattacaa a
caggaatcg
1141  aatgcaaccg gcgcaggaac actgccagcg catcaacaat attttcacct g
aatcaggat
1201  attcttctaa tacctggaat gctgttttcc cggggatcgc agtggtgagt a

```

FIGURE 16M

WO 02/090300

PCT/US02/14358

accatgcac
1261 catcaggagt acggataaaa tgcttgatgg tcggaagagg cataaattcc g
tcagccagc
1321 ttagctctgac catctcatct gtaacatcat tggcaacgct acctttgcca t
gtttcagaa
1381 acaactctgg cgcacgggc ttcccatata atcgatagat tgtcgcacct g
attgcccca
1441 cattatcgcg agccattta taccatata aatcagcatc catgttgaa t
ttaatcgcg
1501 gcctagagca agacgtttcc cgttgaatat ggctcataac accccttgta t
tactgttta
1561 tgtaagcaga cagttttatt gtctatgacc aaaatccctt aacgtgagtt t
tcgttcac
1621 tgagcgtcag acccgtaga aaagatcaa ggatcttctt gagatccctt t
ttctcgcc
1681 gtaatctgct gcttgcaaac aaaaaacca ccgctaccag cgggtggttg t
ttgccgat
1741 caagagctac caactctttt tccgaaggta actggcttca gcagagcgca g
ataccaaat
1801 actgtccttc tagttagacc gtatgtaggc caccacttca agaactctgt a
gcaccgct
1861 acatacctcg ctctgctaact cctgttacca gtggctgctg ccagtgccga t
aagtcgtgt
1921 cttaccgggt tggactcaag acgatagta ccgataagg cgcagcggtc g
ggctgaacg
1981 gggggttcgt gcacacagcc cagcttgag cgaacgacct aaccggaact g
agataccta
2041 cagcgtgagc tatgagaaag cgcacgctt ccgaaggga gaaaggcgga c
aggtatccg
2101 gtaagcggca gggcgggaac aggagagcgc acgagggagc ttccaggggg a
aacgcctgg
2161 tatctttata gtctgtcgg gtttcgccac ctctgacttg agcgtcgatt t
ttgtgatgc
2221 tcgtcagggg ggcggagcct atggaaaaac gccagcaacg cggccttttt a
cggttcctg
2281 gccctttgct ggccttttgc tcacatgttc tttcctgctg tatccctga t
tctgtggat
2341 aaccgtatta ccgcctttga gtgagctgat accgctcgcc gcagccgaac g
accgagcgc
2401 agcaggtcag tgagcgagga agcgggaagag cgcctgatgc ggtattttct c
cttacgcat
2461 ctgtgcggta ttccacaccg catatatggt gcaactctcag tacaatctgc t
ctgatccg
2521 catagttaag ccagtataca ctccgctatc gctacgtgac tgggtcatgg c
tgccgcccg
2581 acaccgccca acaccgctg acgcgccctg acgggcttgt ctgctcccg c
atccgctta
2641 cagacaagct gtgaccgtct ccgggagctg catgtgtcag aggttttcc c

WO 02/090300

PCT/US02/14358

gtcatcacc
2701 gaaacgcgcg aggcagctgc ggtaagctc atcagcgtgg tctgaagcg a
ttcacagat
2761 gtctgcctgt tcatcccggt ccagctcgtt gagtttctcc agaagcgtta a
tgtctggct
2821 tctgataaag cgggccatgt taaggcgggt tttttcctgt ttggtaactg a
tgccctcgt
2881 gtaaggggga tttctgttca tgggggtaat gataccgatg aaacgagaga g
gatgctcac
2941 gatacgggtt actgatgatg aacatgccc gttactggaa cgttgtgagg g
taaaccaact
3001 ggcggtatgg atgcggcggg accagagaaa aatcactcag ggtcaatgcc a
gcgcttcgt
3061 taatacagat gtaggtgttc cacagggtag ccagcagcat cctgcgatgc a
gatccggaa
3121 cataatggtg cagggcgctg acttcgcgtt ttccagactt taacaaacac g
gaaaccgaa
3181 gaccattcat gttgttgctc aggtcgaga cgttttgag cagcagtcgc t
tcacgttcg
3241 ctgcgtatc ggtgattcat tctgctaacc agtaaggcaa ccccgccagc c
tagccgggt
3301 cctcaacgac agggacacga tcatgcgcac ccgtggggcc gccatgccgg c
gataatgac
3361 ctgcttctcg ccgaaacgtt tggtggcggg accagtgaag aaggtttgag c
gagggcgtg
3421 caagattccg aataccgcaa gcgacaggcc gatcatcgtc gcgctccagc g
aaagcggtc
3481 ctgcgcgaaa atgaccaga gcgctgcggg cactgtcct acgagttgca t
gataaagaa
3541 gacagtcata agtgcggcga cgatagtcac gccccgcgcc caccggaagg a
gotgaactg
3601 gttgaaggct ctcaagggca tcggtcgaga tcccggtgcc taatgagtga g
ctaacttac
3661 attaatgag ttgcgctcac tgcccgttt ccagtcggga aacctgttgt g
ccagctgca
3721 ttaatgaatc ggccaacgcg cggggagagg cggtttcgt attgggcgcc a
gggtgggtt
3781 ttcttttcac cagtgaagc ggcaacagct gattgcccct caccgcctgg c
cctgagaga
3841 gttgcagcaa gcggtccacg ctggtttgcc ccagcaggcg aaaatcctgt t
tgatggtgg
3901 ttaacggcgg gatataacat gagctgtctt cggtatcgtc gtatccact a
ccgagatgt
3961 ccgcaccaac gcgcagcccg gactcggtta tggcacgcat tgcgccagc g
ccatctgat
4021 cgttgggaac cagcatcgca gtgggaacga tgccctcatt cagcatttgc a
tggtttgtt
4081 gaaaaccgga catggcactc cagtcgcctt cccgttcgc tatcggtga a

WO 02/090300

PCT/US02/14358

```

tttgattgc
4141 gagttagata tttatgccag ccagccagac gcagacgcgc cgagacagaa c
ttaatgggc
4201 ccgctaacag cgcgatttgc tggtagacca atgcgaccag atgctccacg c
ccagtcgcg
4261 taccgtcttc atgggagaaa ataatactgt tgatgggtgt ctggtcagag a
catcaagaa
4321 ataacgccgg aacattagt caggcagctt ccacagcaat ggcaccttg t
catccagcg
4381 gatagttaat gatcagccca ctgacgcgtt gcgcgagaag attgtgcacc g
ccgctttac
4441 aggcttcgac gccgcttcgt tctaccatcg acaccacacg gctggcacc a
gttgatcgg
4501 cgcgagattt aatcgccgcg acaatttgcg acgcgcgctg caggggccaga c
tggaggtgg
4561 caacgccaat cagcaacgac tgtttgcccg ccagttgttg tgccacgcgg t
tgggaatgt
4621 aattcagctc cgccatcgcc gcttccactt tttcccggt tttcgagaa a
cgtggctgg
4681 cctggttcac cagcgggaa acggtctgat aagagacacc ggcatactct g
cgacatcgt
4741 ataacgttac tggtttcaca ttaccacacc tgaattgact ctcttcggg c
gctatcatg
4801 ccataccgcg aaagggtttg cgccattcga tgggttcgg gatctcgacg c
tctccctta
4861 tgcgactcct gcattagaa gcagcccagt agtaggttga ggccgttgag c
accgcccgc
4921 gcaagggaatg gtgcagcaa ggagatggcg cccaacagtc cccggccac g
gggctgoc
4981 accataccca cgcgaaaca agcgcctatg agcccgaagt ggcgagcccg a
tcttccca
5041 tcggtgatgt cgcgcatata ggccgacga accgcacctg tggcgccggt g
atgcgggcc
5101 acgatgcgtc cgcgtagag gatcgagac gatctcgatc ccgcgaaatt a
atacgactc
5161 actatagggg aattgtgagc ggataacaat tccctctag aaataatttt g
tttaacttt
5221 aagaaggaga tatacatatg aaagaaacg ctgctgctaa attcgaacgc c
agcacatgg
5281 acagcccaga tctgggtacc ctgggtccac gcggttccat ggctgatac a
gatctaattg
5341 gtctcgaaac tcacaacaca aggcctctgta tcgtaggaag tggcccagcg g
cacacacgg
5401 cggcgattta cgcagctagg gctgaactta aacctcttct cttcgaggga t
ggatggcta
5461 acgacatcgc tcccggtggt caactaaca ccaccacga cgtcgagaat t
tccccgat
5521 ttccagaagg tattctcgga gtagagctca ctgacaaatt ccgtaacaa t

```

FIG 16M (CONT)

WO 02/090300

PCT/US02/14358

cggagcgat
5581 tcggtactac gatatttaca gagacgggtga cgaagtgga tttctcttcg a
aaccgttta
5641 agctattcac agattccaag gccattctcg ctgacgctgt gattctogct a
ctggagctg
5701 tggctaagcg gcttagcttc gttggatctg gtgaagggtc tggagggttc t
ggaacggtg
5761 gaatctccgc atgctgtggt tgcgaaggag ctgctccgat attccgtaac a
aacctcttg
5821 cgggtgatcg tggaggcgat tcagcaatgg aagaagcaaa cttctttaca a
aatatggat
5881 ccaaagtgtg tataatccat tggtttgatg cttttcaggc gtctaagatt a
tgcagcagc
5941 gcgctttgtc taatcctaag attgatgtga tttggaactc gtctgttgtg g
aagcttatg
6001 gagatggaga aagagatgtg cttggaggat tgaaagtga gaatgtggtt a
ccggtgatg
6061 tttctgattt aaaagtttct ggattgttct ttgctattgg tcatgagcca g
ctaccaagt
6121 ttttggtatg tgggtgtgag ttagattcgg atggttatgt tgtcacgaag c
ctggtacta
6181 cacagactag cgttccgga gttttcgtg cgggtgatgt tcaggataag a
agtataggc
6241 aagccatcac tgctgcagga actgggtgca tggcagcttt ggatgcagag c
attacttac
6301 aagagattgg atctcagca ggtaagagt attgagtcga caagcttgcg g
ccgcac

FIG 16M (CONT)

WO 02/090300

PCT/US02/14358

The NTR-WT (RRR-WT) coding sequence with S-tag at the N-terminus (5238-6335)

```

1  tctgagcacca ccaccaccac cactgagatc cggctgctaa caaagcccga a
aggaagctg
61  agttggctgc tgccaccgct gagcaataac tagcataacc ctttggggcc t
ctaaacggg
121  tcttgagggg ttttttgcgt aaaggaggaa ctatatccgg attggcgaat g
ggacgcggc
181  ctgtagcggc gcattaagcg cggcgggtgt ggtggttacg cgcagcgtga c
cgctacact
241  tgccagcgcc ctacgcgccg ctcccttcgc tttcttccct tcttttctcg c
cacgttcgc
301  cggctttccc cgtcaagctc taaatcgggg gctcccttta gggttcgat t
tagtgcttt
361  acggcacctc gacccaaaaa aaacttgatta ggggtgatgg tccagtagtg g
gccatcgcc
421  ctgatagacg gtttttcgcc ctttgacgtt ggagtccacg tcttttaata g
tggactctt
481  gttccaaact ggaacaacac tcaaccctat ctcggtctat tcttttgatt t
ataagggat
541  tttgcgattt tcggcctatt ggttaaaaaa tgagctgatt taacaaaaat t
taacgcgaa
601  ttttaacaaa atattaacgt ttacaatttc aggtggcact tttcggggaa a
tgtgcgcgg
661  aacccctatt tgtttatttt tctaaataca ttcaaatatg tatccgctca t
gaattaatt
721  cttagaaaaa ctcacgcgac atcaaatgaa actgcaattt attcatatca g
gattatcaa
781  taccatattt ttgaaaaagc cgtttctgta atgaaggaga aaactcaccg a
ggcagttcc
841  ataggatggc aagatcctgg tatcggtctg cgattccgac tcttccaaca t
caatacaac
901  ctattaattt cccctcgta aaaataaggt tatcaagtga gaaatcacca t
gagtgcgca
961  ctgaatccgg tgagaatggc aaaagtttat gcatttcttt ccagacttgt t
caacaggcc
1021  agccattacg ctcgatcatc aaatcactcg catcaaccaa accgttatto a
ttcgtgatt
1081  gcgcctgagc gagacgaaat acgcgctcgc tgttaaaagg acaattacaa a
caggaatcg
1141  aatgcaaccg gcgcaggaac actgccagcg catcaacaat attttcacct g
aatcaggat
1201  attcttctaa tacctggaat gctgttttcc cggggatcgc agtgggtgag a
accatgcac
1261  catcaggagt acggataaaa tgcttgatgg tcggaagagg cataaattcc g

```

FIGURE 16 N

WO 02/090300

PCT/US02/14358

tcagccagt
1321 ttagttctgac catctcatct gtaacatcat tggcaacgct acctttgcca t
gtttcagaa
1381 acaactcttg cgcctcgggc ttcccatata atcgatagat tgtcgcaoct g
attgccoga
1441 cattatcgcg agcccattta tacocatata aatcagcctc catgttgga t
ttaatcgcg
1501 gcctagagca agacgtttcc cgttgaatat ggctcataac accccttgta t
tactgttta
1561 tgtaagcaga cagttttatt gttcatgacc aaaatccctt aacgtgagtt t
tcgttccac
1621 tgagcgtcag acccgtaga aaagatcaa ggatcttctt gagatccctt t
tttctgcgc
1681 gtaatctgct gcttgcaaac aaaaaacca ccgctaccag cgggtggttg t
tgccggat
1741 caagagctac caactctttt tcogaaggta actggcttca gcagagcgca g
ataccaaat
1801 actgtccttc tagtgtagcc gtagttaggc caccacttca agaactctgt a
gcaccgcct
1861 acatacctcg ctctgctaatt cctgttacca gtggctgctg ccagtggcga t
aagtctgt
1921 cttaccgggt tggactcaag acgtagtga cggataagg cgcagcggtc g
ggctgaacg
1981 gggggttcgt gcacacagcc cagcttgag cgaacgacct acaccgaact g
agataccta
2041 cagcgtgagc tatgagaaag cgcacgcctt cccgaaggga gaaaggcgga c
aggatccg
2101 gtaagcggca gggtcggaac aggagagcgc acgagggagc ttccaggggg a
aacgcctgg
2161 tatctttata gtctgtcgg gtttcgccac ctctgacttg agcgtcgatt t
ttgtgatgc
2221 tcgtcagggg ggcggagcct atggaaaaac gccagcaacg cggccttttt a
cggttcctg
2281 gccttttgcg ggccttttgc tcacatgttc tttcctgcgt tatccctga t
tctgtggat
2341 aaccgtatta ccgcctttga gtgagctgat accgctcgcc gcagccgaac g
accagcgcc
2401 agcaggtcag tgagcgagga agcggaaag cgcctgatgc ggtattttct c
cttacgcat
2461 ctgtgcggta ttccacacgc catatatggt gcactctcag tacaatctgc t
ctgatgcgc
2521 catagttaag ccagtataca ctccgctatc gtaacgtgac tgggtcatgg c
tgccgcccg
2581 acaccgcgca acaccgcctg acgcgcctg acgggcttgt ctgctcccg c
atccgctta
2641 cagacaagct gtgacgctct ccgggagctg catgtgtcag aggttttcac c
gtcatcacc
2701 gaaacgcgcg aggcagctgc ggtaaagctc atcagcgtgg tcgtgaagcg a

FIG 16N (CONT)

WO 02/090300

PCT/US02/14358

```

ttcacagat
2761 gtctgcctgt tcatccgcgt ccagctcgtt gagtttctcc agaagcgtta a
tgtctggct
2821 tctgataaag cgggccatgt taaggcggtt ttttctctgt ttggtcactg a
tgctccgt
2881 gtaaggggga tttctgttca tgggggtaat gataccgatg aaacgagaga g
gatgctoac
2941 gatacgggtt actgatgatg aacatgcccg gttactggaa cgttgtgagg g
taaacaact
3001 ggcggtatgg atgcggcggg accagagaaa aatcactcag ggtcaatgcc a
gcgcttcgt
3061 taatacacat gtaggtgttc cacagggtag ccagcagcat cctgcgatgc a
gatccggaa
3121 cataatggtg caggcgctg acttccgcgt ttccagactt taogaaacac g
gaaaccgaa
3181 gaccattcat gttgttgctc aggtcgaga cgttttgcat cagcagtcgc t
tcacgttcg
3241 ctgcgctatc ggtgattcat tctgctaacc agtaaggcaa cccgcaccgc c
tagccgggt
3301 cctcaacgac aggagcacga tcatgcgcac ccgtggggcc gccatgccgg c
gataatggc
3361 ctgcttctcg cgaacaggtt tgggtggcgg accagtgaag aaggcttgag c
gaggcgctg
3421 caagattccg aataccgcaa gcgacaggcc gatcatcgtc gcgctccagc g
aaagcggtc
3481 ctgcgcgaaa atgaccaga gcgctgcggg cactgtcct acgagttgca t
gataaagaa
3541 gacagtcata agtgcgcgca cgatagtcac gccccgcgcc caccggaagg a
gctgactgg
3601 gttgaaggct ctcaagggca tgggtcgaga tcccggtgcc taatgagtga g
ctaacttac
3661 attaatgctg ttgcgctcac tgcccgctt ccagtcggga aacctgtcgt g
ccagctgca
3721 ttaatgaatc ggccaacgcg cggggagagg cggtttcgt attgggcgcc a
gggtggttt
3781 ttcttttcac cagtgaagac ggcaacagct gattgcctt caccgcctgg c
cctgagaga
3841 gttgcagcaa gcggtccacg ctggtttgcc ccagcaggcg aaaatcctgt t
tgatggtgg
3901 ttaacggcgg gatataacat gagctgtctt cggtatcgtc gtatccact a
ccgagatgt
3961 ccgcaccaac gcgcagccc gactcggtta tggcagcat tgcgccagc g
ccatctgat
4021 cgttggaac cagcatcgca gtgggaacga tgccctcatt cagcatctgc a
tggtttgtt
4081 gaaaaccgga catggcactc cagtcgcctt ccggttcgc tatcggtga a
tttgattgc
4141 gagtgaata tttatgccag ccagccagac gcagacgcgc cgagacagaa c

```

FIG 16N (CONT)

WO 02/090300

PCT/US02/14358

```

ttaatgggc
4201 ccgctaacag cgcgatttgc tggtagccca atgcgaccag atgctccacg c
ccagtgcgc
4261 taccgtcttc atgggagaaa ataatactgt tgatgggtgt ctggtcagag a
catcaagaa
4321 ataacgcgcg aacattagtg caggcagctt ccacagcaat ggcctcttg t
catccagcg
4381 gatagttaat gatcagccca ctgacgcgtt gcgcgagaag attgtgcacc g
ccgctttac
4441 aggcttcgac gccgcttctg totaccatcg acaccaccac gctggcaccd a
gttgatcgg
4501 cgcgagattt aatgcgcgcg acaatttgcg acggcgcgtg cagggccaga c
tggaggtgg
4561 caacgccaat cagcaacgac tgtttgcccg ccagttgttg tgccacgcgg t
tgggaatgt
4621 aattcagctc cgccatcgcc gcttccactt ttcccgctg ttctgcagaa a
cgtggctgg
4681 cctgggtcac cgcgcgggaa acggtctgat aagagacacc ggcatactct g
cgacatcgt
4741 ataacgttac tggtttcaca ttcaccaccc tgaattgact ctcttcggg c
gctatcatg
4801 ccataccgcg aaagggtttg cgccattcga tgggttcggg gatctcgacg c
tctccctta
4861 tgcgactcct gcattaggaa gcagcccagt agtaggttga ggcggttgag c
accgcgcgc
4921 gcaaggaatg gtgcctgcaa ggagatggcg cccaacagtc ccccgccac g
gggcctgcc
4981 accataccca cgcgaaaca agcgcctatg agcccgagt ggcgagcccg a
tcttcccca
5041 tcggtgatgt cggcgatata ggcgcacga accgcacctg tggcgccggt g
atgcgcgcg
5101 acgatgcgtc cggcgtagag gatcgagtc gatctgac cgcgaaatt a
atacgaact
5161 actatagggg aattgtgagc ggataacaat tcccctctag aaataatttt g
tttaacttt
5221 aagaaggaga tatacatatg aaagaaaccg ctgctgctaa attcgaaacg c
agcacatgg
5281 acagcccaga tctgggtacc ctggtgcccac gcggttccat ggcgtgatatc a
gatctaattg
5341 gtctcgaaac tcacaacaca aggtctctga tcgtaggaag tggcccagcg g
cacacacgg
5401 cggcgattta cgcagctagg gctgaactta aacctcttct ctctgaagga t
ggatggcta
5461 acgacatcgc tcccgtgggt caactaaca caccaccga cgtcgagaat t
tcccggat
5521 ttccagaagg tattctcgga gttagctca ctgacaaatt ccgtaaaca t
cggagcgat
5581 tcggtactac gatatttaca gagacggtga cgaaagtcga tttctcttcg a

```

FIG 16N (CONT)

WO 02/090300

PCT/US02/14358

aacogttta
5641 agctattcac agattccaag gccattctcg ctgacgctgt gattctcgct a
ctggagctg
5701 tggctaagcg gcttagcttc gttgcatctg gtgaagggttc tggagggttc t
ggaaccgtg
5761 gaatctccgc atgcgtgttt tgcgacggag ctgctccgat attccgtaac a
aacctcttg
5821 cggatgacgg tggaggcgat tcagcaatgg aagaagcaaa ctttcttaca a
aatatggat
5881 ccaaagtgtg tataatccat aggagagatg cttttagagc gtctaagatt a
tgcacagc
5941 gcgctttgtc taatcctaag attgatgtga ttggaactc gtctgttggtg g
aagcttatg
6001 gagatggaga aagagatgtg cttggaggat tgaagtga gaatgtggtt a
cgggtgatg
6061 tttctgattt aaaagtttct ggattgttct ttgctattgg tcatgagcca g
ctaccaagt
6121 ttttggatgg tgggtgtgag ttagattcgg atggttatgt tgtcacgaag c
ctggtacta
6181 cacagactag cgttccgga gttttcgctg cgggtgatgt tcaggataag a
agtataggc
6241 aagccatcac tgctgcagga actgggtgca tggcagcttt ggatgcagag c
attaottac
6301 aagagattgg atctcagcaa ggtaagagtg attgagtoga caagcttgcg g
ccgcac

FIG 16N (CONT)

WO 02/090300

PCT/US02/14358

The RYN-M variant coding sequence with S-tag at the N-terminus (5238-6335)

```

1  tcgagcacca ccaccaccac cactgagatc cggctgctaa caaagcccga a
aggaagctg
61  agttggctgc tgccaccgct gagcaataac tagcataacc ccttggggcc t
ctaaacggg
121  tcttgagggg ttttttgctg aaaggaggaa ctatatccgg attggcgaat g
ggacgcgcc
181  ctgtagcgcc gcattaagcg cggcgggtgt ggtgggttacg cgcagcgtga c
cgctacact
241  tgccagcgcc ctacgcgcc ctcctttcgc tttcttccct tectttctcg c
cacgttcgc
301  cggctttccc cgtcaagctc taaatcgggg gctcccttta gggttccgat t
tagtgcttt
361  accggcacctc gaccccaaaa aacttgatta gggatgatgtg tcacgtagtg g
gccatcgcc
421  ctgatagacg gtttttcgcc ctttgacgtt ggagtccacg ttttttaata g
tggactctt
481  gttccaaact ggaacaacac tcaaccctat ctgggtctat tcttttgatt t
ataagggat
541  tttgccgatt tcggcctatt ggttaaaaaa tgagctgatt taacaaaaat t
taacgcgaa
601  ttttaacaaa atattaacgt ttacaatttc aggtggcact ttcgggggaa a
tgtgcgcgg
661  aaccctctatt tgtttatatt tctaaatata ttcaaatatg tatccgctca t
gaattaatt
721  cttagaaaaa ctcacgagc atcaaatgaa actgcaattt attcatatca g
gattatcaa
781  taccatattt ttgaaaaago cgtttctgta atgaaggaga aaactcacgc a
ggcagttcc
841  ataggatggc aagatcctg tctcggctcg cgtttccgac tcgtccaaca t
caatacaac
901  ctattaattt cccctcgtca aaaataaggt tatcaagtga gaaatcacca t
gagtgcgca
961  ctgaatcccg tgagaatggc aaaagtttat gcattttctt ccagacttgt t
caacaggcc
1021  agccattacg ctgctcatca aaatcactcg catcaaccaa accgttattc a
ttcgtgatt
1081  gcgcctgagc gagacgaaat acgcgatcgc tgttaaaagg acaattacaa a
caggaatcg
1141  aatgcaaccc gcgcaggaac actgccagcg catcaacaat atttcacct g
aatcaggat
1201  attcttctaa tacctggaat gctgttttcc cggggatcgc agtgggtgagt a
accatgcat
1261  catcaggagt acggataaaa tgcttgatgg tcggaagagg cataaattcc g

```

FIGURE 16 O

WO 02/090300

PCT/US02/14358

```

tcagccagt
1321 ttagctctgac catctcatct gtaacatcat tggcaacgct acctttgcca t
gtttcagaa
1381 acaactctgg cgcctcgggc ttcccataca atcgatagat tgtegcacct g
attgccga
1441 cattatcgcg agcccattta taccatata aatcagcadc catgttggaa t
ttaatcgcg
1501 gcctagagca agacgtttcc cgttgaatat ggctcataac accccttgta t
tactgttta
1561 tgtaagcaga cagttttatt gttcatgacc aaaatccott aacgtgagtt t
tcgttccac
1621 tgagcgtcag acccctaga aaagatcaaa ggatctctt gagatccttt t
ttctgcgc
1681 gtaatctgct gcttgcaaac aaaaaaacca ccgctaccag cgggtggtttg t
ttgcggat
1741 caagagctac caactctttt tccgaaggta actggcttca gcagagcgca g
ataccaaat
1801 actgtccttc tagtgtagcc gtagttaggc caccacttca agaactctgt a
gcaccgcct
1861 acatacctcg ctctgctaatt cctgttacca gtggctgctg ccagtggcga t
aagtcgtgt
1921 cttaccgggt tggactcaag acgatagtta ccgataaagg cgcagcggtc g
ggctgaacg
1981 ggggggttcgt gcacacagcc cagcttggag cgaacgacct acaccgaact g
agataccta
2041 cagcgtgagc tatgagaaag cgcacgcctt cccgaaggga gaaaggcgga c
aggtatccg
2101 gtaagcgga gggtcggaac aggagagcgc acgagggagc ttccaggggg a
aacgcctgg
2161 tatctttata gtcctgtcgg gtttcgccac ctctgacttg agcgtcgatt t
ttgtgatgc
2221 tcgtcagggg ggcggagcct atggaaaaac gccagcaacg cggccttttt a
cggttcctg
2281 gccttttgct ggccttttgc tcacatgttc ttctcgtct tatccctga t
tctgtggat
2341 aaccgtatta ccgcctttga gtgagctgat accgtcggc gcagccgaac g
accgagcgc
2401 agcagagtcag tgagcgagga agcgggaagag cgcctgatgc ggtattttct c
cttacgcct
2461 ctgtgcggta ttccacaccg catatatggt gcaactctcag tacaactctgc t
ctgatgccg
2521 catagttaag ccagtataca ctccgctatc gctacgtgac tgggtcatgg c
tgcgcccg
2581 acaccgcga acaccgctg acgcgccctg acgggcttgt ctgctcccg c
atccgcta
2641 cagacaagct gtgacgtct cggggagctg catgtgtcag aggttttcac c
gtcatcac
2701 gaaacgcgag aggcagctgc ggtaaagctc atcagcgtgg tcgtgaagcg a

```

FIG 16 0 (CONT)

WO 02/090300

PCT/US02/14358

```

ttcacagat
2761 gtctgcctgt tcatccgcgt ccagctcgtt gagtttctcc agaagcgta a
tgtctggct
2821 tctgataaag cgggccatgt taaggcggtt ttttctctgt ttggtcactg a
tgctccgt
2881 gtaaggggga tttctgttca tgggggtaat gataccgatg aaacgagaga g
gatgctcac
2941 gatacgggtt actgatgatg aacatgcccg gttactggaa cggtgtgagg g
taaacaact
3001 ggcggtatgg atcgcgcggt accagagaaa aatcaactcag ggtcaatgcc a
gcgcttcgt
3061 taatacagat gtaggtgttc cacagggtag ccagcagcat cctgcgatgc a
gatccgaa
3121 cataatgggtg caggggcgctg acttcgcgtt tccagaactt tacgaaacac g
gaaaccgaa
3181 gaccattcat gttgttgctc aggtcgacga cgttttgcag cagcagtcgc t
tcacgttcg
3241 ctgcgctatc ggtgattcat tctgctaacc agtaaggcaa ccccgccagc c
tagccgggt
3301 cctcaacgac aggagcacga tcatgcgcac ccgtggggcc gccatgccgg c
gataatggc
3361 ctgcttctcg ccgaaacgtt tggtagcggtg accagtgcag aaggcttgag c
gaggcgctg
3421 caagattccg aataaccgaa gcgacaggcc gatcatcgto gcgctccagc g
aaagcggtc
3481 ctgcgcgaaa atgaccaga gcgctgccgg cacctgtcct acgagttgca t
gataaagaa
3541 gacagtcata agtgcgcgga cgatagtcac gcccgcgcgc caccggaagg a
gctgactgg
3601 gttgaaggct ctcaagggca tcggctcgaga tcccggtgcc taatgagtga g
ctaacttac
3661 attaatgtcg ttgcgctcac tgcgcgtttt ccagtcggga aacctgtcgt g
ccagctgca
3721 ttaatgaatc ggccaacgcg cggggagagg cggtttgcgt attgggcgc c
gggtggttt
3781 ttcttttcac cagtgcagcg ggcaacagct gattgccctt caccgcctgg c
cctgagaga
3841 gttgcagcaa gcggtccacg ctggtttgcc ccagcaggcg aaaatcctgt t
tgatggtgg
3901 ttaacggcgg gatataacat gagctgtctt cggtatcgtc gtatccact a
ccgagatgt
3961 ccgcaccaac gcgcagcccg gactcggtaa tggcacgcat tgcgcccagc g
ccatctgat
4021 cggtggcaac cagcatcgca gtgggaacga tgccctcatt cagcatttgc a
tggtttggt
4081 gaaaaccgga catggcactc cagtcgcctt cccgttcgcg tatcggtga a
tttgattgc
4141 gagtgcagata tttatgccag ccagccagac gcagacgcgc cgagacagaa c

```

FIG 16 0 (CONT)

97/113

WO 02/090300

PCT/US02/14358

```

ttaatgggc
4201 ccgctaacag cgcgatttgc tggtagacca atgcgaccag atgctccacg c
coagtccgc
4261 taccgtcttc atgggagaaa ataatactgt tgatgggtgt ctggtcagag a
catcaagaa
4321 ataacgccgg aacattagtg caggcagctt ccacagcaat ggcatacctg t
catccagcg
4381 gatagttaat gatcagccca ctgacgcgtt gcgcgagaag attgtgcacc g
ccgctttac
4441 aggccttcgac gccgcttcgt tctaccatcg acaccaccac gctggcacc a
gttgatcgg
4501 cgcgagattt aatcgccgcg acaatttgcg acggcgcgtg cagggccaga c
tggaggtgg
4561 caacgccaat cagcaacgac tgtttgcccg ccagttgttg tgccacgcgg t
tgggaatgt
4621 aattcagctc cgccatcgcc gcttccactt ttcccgctg ttccgcagaa a
cgtggctgg
4681 cctggttcac cgcggggaa acggtctgat aagagacacc ggcatactct g
cgacatcgt
4741 ataacgttac tggtttcaca ttcaccacc tgaattgact ctcttccggg c
gctatcatg
4801 ccataccgcg aaagggtttg cgccattcga tgggtgccgg gatctcgacg c
tctccctta
4861 tgcgactcct gcattaggaa gcagcccgat agtaggttga ggcggttgag c
accgcccgc
4921 gcaaggaatg gtgcattgca ggagatggcg cccaacagtc ccccgccac g
gggcctgcc
4981 accataccca cgcgaaaca agcgcctcatg agccggaagt ggcgagcccg a
tcttcccca
5041 tcggtgatgt cggcgatata ggcgccagca accgcacctg tggcgccggg g
atgcgggcc
5101 acgatgcgtc cggcgtagag gatcgagatc gatctcgatc ccgcgaaatt a
atacgactc
5161 actatagggg aattgtgagc ggataacaat tccctcttag aaataatttt g
tttaacttt
5221 aagaaggaga tatacatatg aaagaaaccg ctgctgctaa attcgaacgc c
agcacatgg
5281 acagcccaga tctgggtacc ctgggtccac gcggttccat ggctgatata a
gatctaatt
5341 gtctcgaaac tcacaacaca aggctctgta tcgtaggaag tggcccgagc g
cacacacgg
5401 cggcgattta cgcagctagg gctgaactta aacctctct ctccgaagg t
ggatggcta
5461 acgacatcgc tcccggtggg caactaaca ccaccaccga cgtcgagaat t
tccccggat
5521 ttccagaagg tattctcgga gtagagctca ctgacaaatt ccgtaacaa t
cggagcgat
5581 tcggtactac gatatttaca gagacggtga cgaaagtcga tttctcttcg a

```

FIG 16 0 (CONT)

WO 02/090300

PCT/US02/14358

aacgcgttta
5641 agctattcac agattccaag gccattctcg ctgacgctgt gattctcgct a
ctggagctg
5701 tggctaagcg gcttagcttc gttggatctg gtgaaggctc tggagggttc t
ggaaccgtg
5761 gaatctccgc atgcgctgtt tgcgacggag ctgctccgat attccgtaac a
aacctcttg
5821 cggatgatggg tggaggcgat tcagcaatgg aagaagcaaa ctttcttaca a
aatatggat
5881 ccaaagtgtg tataatccat cgctacgatg cttttaatgc gtctaagatt a
tgcagcagc
5941 gcgctttgtc taatcctaag attgatgtga tttggaactc gtctgttggtg g
aagcttatg
6001 gagatggaga aagagatgtg cttggaggat tgaagtgaa gaatgtgggt a
cgggtgatg
6061 tttctgattt aaaagtttct ggattgttct ttgctattgg tcatgagcca g
ctaccaagt
6121 ttttggatgg tgggtgttag ttagattcgg atggttatgt tgtcacgaag c
ctggtaacta
6181 cacagactag cgttcccgga gttttcgtg cgggtgatgt tcaggataag a
agtataggc
6241 aagccatcac tgcgcagga actgggtgca tggcagcttt ggatgcagag c
attacttac
6301 aagagattgg atctcagcaa ggtaagagtg attgagtcga caagcttgcg g
ccgcac

FIG 16 0 (CONT)

WO 02/090300

PCT/US02/14358

The RYN-L variant coding sequence with S-tag at the N-terminus (5238-6335)

```

1  tcgagcacca ccaccaccac cactgagatc cggctgctaa caaagcccga a
aggaagctg
61  agttggctgc tgccaccgct gagcaataac tagcataacc ccttggggcc t
ctaaacggg
121  tcttgagggg ttttttgcgt aaaggaggaa ctatatccgg attggcgaat g
ggacgcgcc
181  ctgtagcggc gcattaagcg cggcgggtgt ggtggttacg cgcagcgtga c
cgctacact
241  tgccagcgcc ctagcgcccg ctcttttcgc tttcttcctt tcctttctcg c
caogttcgc
301  cggctttccc cgtcaagctc taaatcgggg gctcccttta gggttccgat t
tagtgcttt
361  acggcacctc gaccccaaaa aacttgatta gggtagtggt tcacgtagt g
gccatcgcc
421  ctgatatagc gtttttcgcc ctttgacgtt ggagtccacg tttttaata g
tggactctt
481  gttccaaact ggaacaacac tcaaccctat ctcggtctat tcttttgatt t
ataagggat
541  ttgcccattt tcggcctatt ggttaaaaaa tgagctgatt taacaaaaat t
taacgcgaa
601  ttttaacaaa atattaacgt ttacaatttc aggtggcact tttcggggaa a
tgtgcggg
661  aacccctatt tgtttatttt tctaaataca ttcaaatacg tatccgctca t
gaattaatt
721  cttagaaaaa ctcatcgagc atcaaatgaa actgcaattt attcatatca g
gattatcaa
781  taccatattt ttgaaaaagc cgtttctgta atgaaggaga aaactcaccg a
ggcagttcc
841  ataggatggc aagatcctgg tatcggtctg cgattccgac tcgtccaaca t
caatacaac
901  ctattaattt cccctcgtea aaaataaggt tatcaagtga gaaatcacca t
gagtgcgga
961  ctgaatccgg tgagaatggc aaaagtttat gcatttcttt ccagacttgt t
caacaggcc
1021  agccattacg ctogtcatca aaatcactcg catoaaccaa accgttattc a
ttogtgatt
1081  gcgcctgagc gagacgaaat acgcgatcgc tgttaaaagg acaattacaa a
caggaatcg
1141  aatgcaaccg gcgcaggaa cctgccagcg catoaacaat attttcacct g
aatcaggat
1201  attcttctaa tacctggaat gctgttttcc cggggatcgc agtggtgagt a
accatgcat
1261  catcaggagt acggataaaa tgcttgatgg tcggaagagg cataaattcc g

```

FIGURE 16P

WO 02/090300

PCT/US02/14358

```

tcagccagt
1321 ttagtctgac catctcatct gtaacatcat tggcaacgct acctttgcca t
gtttcagaa
1381 acaactctgg cgcctcgggc ttccataca atcgatagat tgtcgacact g
attgcccga
1441 cattatcgcg agcccattta taccatata aatcagcato catgttgaa t
ttaatcgcg
1501 gcctagagca agacgtttcc cgttgaatat ggctcataac accccttgta t
tactgttta
1561 tgtaagcaga cagttttatt gtctatgacc aaaatccctt aacgtgagtt t
tcgttccac
1621 tgagcgtcag accccgtaga aaagatcaaa ggatcttctt gagatccttt t
tttctcgcg
1681 gtaatctgct gcttgcaaac aaaaaaacca ccgctaocag cggtggtttg t
ttgcggat
1741 caagagctac caactctttt tccgaaggta actggcttca gcagagcgca g
ataccaaat
1801 actgtccttc tagtgtagcc gtagttaggc caccacttca agaactctgt a
gcaccgcct
1861 acatacctcg ctctgctaatt cctgttacca gtggctgctg ccagtggcga t
aagtctgt
1921 ctaccgggtg tggactcaag acgatagtta ccggataagg cgcagcggtc g
ggctgaacg
1981 gggggttcgt gcacacagcc cagcttgag cgaacgacct acaccgaact g
agataccta
2041 cagcgtgagc tatgagaaa cgcacgcctt cccgaaggga gaaaggcgga c
aggatatccg
2101 gtaagcggca gggtcggaac aggagagcgc acgagggagc ttccaggggg a
aacgcctgg
2161 tatctttata gtctgtcgg gtttcgccac ctctgacttg agcgtcgatt t
ttgtgatgc
2221 tcgtcagggg ggcggagcct atgaaaaaac gccagcaacg cggccttttt a
cggttcctg
2281 gccttttgct ggccttttgc tcacatgttc tttcctgctg tatccctga t
tctgtggat
2341 aaccgtatta ccgcctttga gtgagctgat accgctcgcc gcagccgaac g
accgagcgc
2401 agcaggtcag tgagcgagga agcgggaagag cgcctgatgc ggtattttct c
cttacgcat
2461 ctgtgcggta ttccacacc catatatggt gcactctca gacaatctgc t
ctgatgccg
2521 catagttaag ccagtataca ctccgctatc gctacgtgac tgggtcatgg c
tgcgccccg
2581 acaccgccca acaccgctg acgcgccctg acgggcttgt ctgctcccg c
atccgctta
2641 cagacaagct gtgaccgtct ccgggagctg catgtgtcag aggttttcac c
gtcatcacc
2701 gaaacgcgcg aggcagctgc ggtaagctc atcagcgtgg tcgtgaagcg a

```

FIG 16P (CONT)

101/113

WO 02/090300

PCT/US02/14358

```

ttcacagat
2761 gtctgcctgt tcattccgct ccagctcgtt gagtttctcc agaagcgtta a
tgtctggct
2821 tctgataaag cgggccatgt taaggcggt ttttctctgt ttggtoactg a
tgctccgt
2881 gtaaggggga tttctgttca tgggggtaat gataccgatg aaacgagaga g
gatgctcac
2941 gatacgggtt actgatgatg aacatgcccg gttactggaa cgttgtgagg g
taaacaaact
3001 ggcggtatgg atgcggcggt accagagaaa aatcactcag ggtcaatgcc a
gcgcttcgt
3061 taatacagat gtaggtgttc cacagggtag ccagcagcat cctgcgatgc a
gatccggaa
3121 cataatgggt caggcgctg acttccgctt ttccagactt tacgaaacac g
gaaaccgaa
3181 gaccattcat gttgttgctc aggtcgaga cgttttgag cagcagtcgc t
tcacgttcg
3241 ctccgctatc ggtgattcat tctgctaacc agtaaggcaa cccgccagc c
tagccgggt
3301 cctcaacgac aggagcacga tcctgcgcac ccgtggggcc gccatgcgg c
gataatggc
3361 ctgcttctcg ccgaaacgtt tggggcggt accagtgcag aaggcttgag c
gagggcggt
3421 caagattccg aataccgcaa gcgacaggcc gatcatcgtc gcgctccagc g
aaagcggtc
3481 ctgcgcgaaa atgaccaga gcgctgcgg cacctgtcct acgagttgca t
gataaagaa
3541 gacagtcata agtcggcgca ctagtagcat gccccgcgac caccggaagg a
gctgactgg
3601 gttgaaggct ctcaagggca tcggtcgaga tcccggtgcc taatgagtga g
ctaaccttac
3661 attaatgctg ttgcgctcac tgcccgctt ccagtcggga aacctgtcgt g
ccagctgca
3721 ttaatgaatc ggccaacgct cggggagagg cgttttgct attgggcgcc a
gggtgggtt
3781 ttcttttca cagtgcagc ggcaacagct gattgcctt caccgcctgg c
cctgagaga
3841 gttgcagcaa gcggtccaog ctggtttgco ccagcaggcg aaaatcctgt t
tgatggtgg
3901 ttaacggcgg gatataacat gagctgtctt cggtatcgtc gtatccact a
ccgagatgt
3961 ccgcaccaac gcgcagccc gactcggtta tggcagcat tgcgcacagc g
ccatctgat
4021 cgttggaac cagcatcgca gtgggaacga tgccctcatt cagcatttgc a
tggtttgtt
4081 gaaaaccgga catggcactc cagtcgcctt cccgttcgc tatcggtga a
tttgattgc
4141 gagtgcagata tttatgccag ccagccagac gcagcgcgc cgagacagaa -c

```

FIG 16P (CONT)

WO 02/090300

PCT/US02/14358

ttaatgggc
4201 ccgctaacag cgcgatttgc tggtagacca atgcgaccag atgctccacg c
ccagtcgcg
4261 tacogtcttc atgggagaaa ataatactgt tgatgggtgt ctggtcagag a
catcaagaa
4321 ataacgc'cgg aacattagt caggcagctt ccacagcaat ggcac'cctg t
catccagcg
4381 gatagttaat gatcagccca ctgacgcgtt gcgcgagaag attgtgcacc g
ccgctttac
4441 aggcttcgac gccgcttcgt tctaccatcg acaccaccac gctggcacc a
gttgatcgg
4501 cgcgagattt aatcgccgcg acaatttgcg acggcgcggtg cagggccaga c
tggaggtgg
4561 caacgccaat cagcaacgac tgtttgcccg ccagttgttg tgccacgcgg t
tgggaatgt
4621 aattcagctc cgcctatgcc gcttccactt tttcccgct tttcgagaa a
cgtggctgg
4681 cctgtttcac cagcgggaa acggtctgat aagagacaco ggcatactct g
cgacatcgt
4741 ataacgttac tggtttcaca ttcaccacc tgaattgact ctcttcggg c
gctatcatg
4801 ccataccgcg aaagggtttg cgcctatcga tgggtgcgg gatctcgacg c
tctccctta
4861 tgcgactcct gcattaggaa gcagcccagt agtaggttga ggcggttgag c
accgcccgc
4921 gcaaggaatg gtgcatgcaa ggagatggcg cccaacagtc cccggccac g
gggctgccc
4981 accataccca cgcgaaaca agcgctcatg agcccgaagt ggcgagccc a
tcttccca
5041 tgggtgatgt cggcgatata ggcgccagca accgcacctg tggcgccggt g
atgccggcc
5101 acgatgcgtc cggcgtagag gatcgagatc gatctcgatc ccgcaaat a
atacgactc
5161 actatagggg aattgtgagc ggataacaat tcccctctag aaataat'ctt g
t'ctaa'ctt
5221 aagaaggaga tatacatatg aaagaaacgc ctgctgctaa attogaacgc c
agcacatgg
5281 acagccaga tctgggtacc ctggtgccac gcggttccat ggtgatato a
gatctaag
5341 gtctcgaaac tcacaacaca aggtctgtga tctaggaag tggcccagcg g
cacacagg
5401 cggcgattta cgcagctagg gctgaactta aacctcttct ctctgaagga t
ggatggcta
5461 acgacatcgc tcccgtgggt caactaacia cccaccaga cgtcgagaat t
tcccggat
5521 ttccagaagg tattctcgga gtagagctca ctgacaaatt ccgtaacia t
cggagcgat
5581 tcggtactac gatatttaca gagacggtga cgaaagtcga tttctcttcg a

FIG 16P (CONT)

WO 02/090300

PCT/US02/14358

aaccgttta
5641 agctattcac agattccaag gccattctcg ctgacgctgt gattctcgct a
ctggagctg
5701 tggctaagcg gcttagcttc gttggatctg gtgaaggttc tggagggttc t
ggaaccgtg
5761 gaatctccgc atgcgctgtt tgcgacggag ctgctccgat attccgtaac a
aacctcttg
5821 cggctgctggg tggaggcgat tcagcaatgg aagaagcaaa ctttcttaca a
aatatggat
5881 ccaaagtgtataataatccat cgctacgatg cttttaatgc gtctaagatt a
tgcagcagc
5941 gcgctttgtc taatcctaag attgatgtga tttggaactc gtctgttgtg g
aagcttatg
6001 gagatggaga aagagatgtg cttggaggat tgaaagtga gaatgtggtt a
cgggtgatg
6061 tttctgattt aaaagtttct ggattgttct ttgctattgg tcatgagcca g
ctaccaagt
6121 ttttggatgg tgggtgtgag ttagattcgg atgggttatgt tgtcacgaag c
ctggtacta
6181 cacagactag cgttcccga gttttcgctg cgggtgatgt tcaggataag a
agtataggc
6241 aagccatcac tgctgcagga actgggtgca tggcagcttt ggatgcagag c
attacttac
6301 aagagattgg atctcagcaa ggtaagagtg attgagtcga caagcttgcg g
ccgcac

FIG 16P (CONT)

104/113

WO 02/090300

PCT/US02/14358

Alignment of NTR WT and new variant protein sequences, as encoded by the expression vector used. Please refer to color legend below for description of amino acid positions.

[illegible][illegible][illegible][illegible]

LEGEND:

[illegible]

Non-ERR Positions Designed in TR-2 Library

© 2004 by The McGraw-Hill Companies, Inc.

FIGURE 17A

105/113

WO 02/090300

PCT/US02/14358

Alignment of NTR WT and variant protein sequences, as encoded by the expression vector used. The mutations are indicated in blue/bold. The cleavage site for thrombin used in S-tag purification strategy is underlined and indicated by red arrow.

FileUp

```

MSF: 365  Type: P  Check: 750  ..

Name: RRR-WT-S-Tagged_Protein  Len: 365  Check: 37  Weight: 0
Name: RYN-I-S-Tagged_Protein  Len: 365  Check: 177  Weight: 0
Name: RYN-L-S-Tagged_Protein  Len: 365  Check: 255  Weight: 0
Name: RYN-M-S-Tagged_Protein  Len: 365  Check: 281  Weight: 0

//

1                               50
RRR-WT-S-Tagged_Protein  MKETAAAKFE  RQHMDSPDLG  TLVPRGSMAD  IRSNGLETHN  TRLCIVGSGP
RYN-I-S-Tagged_Protein  MKETAAAKFE  RQHMDSPDLG  TLVPRGSMAD  IRSNGLETHN  TRLCIVGSGP
RYN-L-S-Tagged_Protein  MKETAAAKFE  RQHMDSPDLG  TLVPRGSMAD  IRSNGLETHN  TRLCIVGSGP
RYN-M-S-Tagged_Protein  MKETAAAKFE  RQHMDSPDLG  TLVPRGSMAD  IRSNGLETHN  TRLCIVGSGP
                               ↑

51                               100
RRR-WT-S-Tagged_Protein  AAHTAAIYAA  RAEIKFLLPE  GWMANDIAPG  GQLTTTTDVE  NFPGFPEGIL
RYN-I-S-Tagged_Protein  AAHTAAIYAA  RAEIKFLLPE  GWMANDIAPG  GQLTTTTDVE  NFPGFPEGIL
RYN-L-S-Tagged_Protein  AAHTAAIYAA  RAEIKFLLPE  GWMANDIAPG  GQLTTTTDVE  NFPGFPEGIL
RYN-M-S-Tagged_Protein  AAHTAAIYAA  RAEIKFLLPE  GWMANDIAPG  GQLTTTTDVE  NFPGFPEGIL

101                              150
RRR-WT-S-Tagged_Protein  GVELTDKFRK  QSERFGTTIF  TETVTKVDFS  SKPFKLFIDS  KAILADAVIL
RYN-I-S-Tagged_Protein  GVELTDKFRK  QSERFGTTIF  TETVTKVDFS  SKPFKLFIDS  KAILADAVIL
RYN-L-S-Tagged_Protein  GVELTDKFRK  QSERFGTTIF  TETVTKVDFS  SKPFKLFIDS  KAILADAVIL
RYN-M-S-Tagged_Protein  GVELTDKFRK  QSERFGTTIF  TETVTKVDFS  SKPFKLFIDS  KAILADAVIL

151                              200
RRR-WT-S-Tagged_Protein  ATGAVAKRLS  FVSGSGSGSG  FWNRGISACA  VCDGAAPIFR  NKPLAVIGGG
RYN-I-S-Tagged_Protein  ATGAVAKRLS  FVSGSGSGSG  FWNRGISACA  VCDGAAPIFR  NKPLAVIGGG
RYN-L-S-Tagged_Protein  ATGAVAKRLS  FVSGSGSGSG  FWNRGISACA  VCDGAAPIFR  NKPLAVIGGG
RYN-M-S-Tagged_Protein  ATGAVAKRLS  FVSGSGSGSG  FWNRGISACA  VCDGAAPIFR  NKPLAVIGGG

201                              250
RRR-WT-S-Tagged_Protein  DSAMBEANFL  TKYGSKVYII  HRYDAPNASK  IMQORALSNP  KIDVINNSV
RYN-I-S-Tagged_Protein  DSAMBEANFL  TKYGSKVYII  HRYDAPNASK  IMQORALSNP  KIDVINNSV
RYN-L-S-Tagged_Protein  DSAMBEANFL  TKYGSKVYII  HRYDAPNASK  IMQORALSNP  KIDVINNSV
RYN-M-S-Tagged_Protein  DSAMBEANFL  TKYGSKVYII  HRYDAPNASK  IMQORALSNP  KIDVINNSV

251                              300
RRR-WT-S-Tagged_Protein  VEAYGDGERD  VLGLKVKNV  VTGDVSDLKV  SGLFFAIGHE  PATKFLDGGV
RYN-I-S-Tagged_Protein  VEAYGDGERD  VLGLKVKNV  VTGDVSDLKV  SGLFFAIGHE  PATKFLDGGV
RYN-L-S-Tagged_Protein  VEAYGDGERD  VLGLKVKNV  VTGDVSDLKV  SGLFFAIGHE  PATKFLDGGV
RYN-M-S-Tagged_Protein  VEAYGDGERD  VLGLKVKNV  VTGDVSDLKV  SGLFFAIGHE  PATKFLDGGV

301                              350
RRR-WT-S-Tagged_Protein  ELDSGYYVVT  KPGTTQTSVP  GVFAAGDVQD  KKYRQAITAA  GTGCMALDA
RYN-I-S-Tagged_Protein  ELDSGYYVVT  KPGTTQTSVP  GVFAAGDVQD  KKYRQAITAA  GTGCMALDA
RYN-L-S-Tagged_Protein  ELDSGYYVVT  KPGTTQTSVP  GVFAAGDVQD  KKYRQAITAA  GTGCMALDA
RYN-M-S-Tagged_Protein  ELDSGYYVVT  KPGTTQTSVP  GVFAAGDVQD  KKYRQAITAA  GTGCMALDA

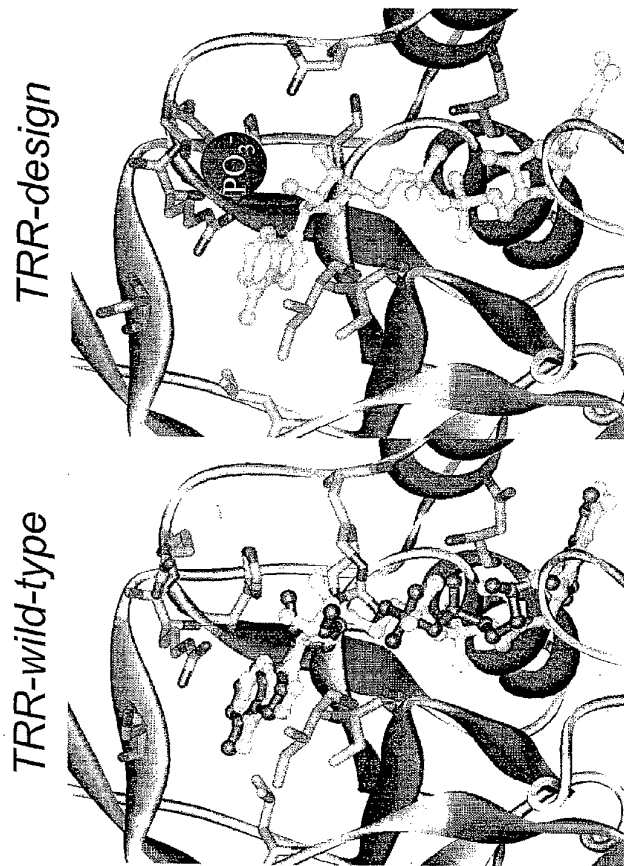
351                              365
RRR-WT-S-Tagged_Protein  EHYLQEIGSQ  QGKSD
RYN-I-S-Tagged_Protein  EHYLQEIGSQ  QGKSD
RYN-L-S-Tagged_Protein  EHYLQEIGSQ  QGKSD
RYN-M-S-Tagged_Protein  EHYLQEIGSQ  QGKSD

```

FIGURE 17B
106/113

WO 02/090300

PCT/US02/14358

**FIGURE 18**

WO 02/090300

PCT/US02/14358

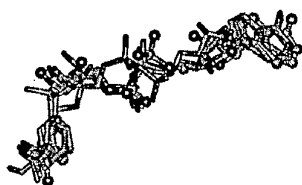


FIGURE 19

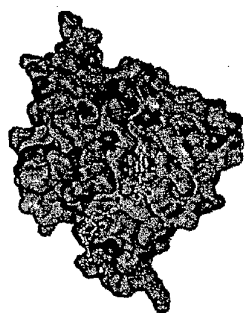


FIGURE 20

WO 02/090300

PCT/US02/14358

Figure 21A

>sp|P09625|TRXB_ECOLI Thioredoxin reductase (EC 1.6.4.5) (TRXR) - Escherichia coli, and Escherichia coli O157:H7.
 GFTKHSKLLIGSSRGATTANTYAAANLQPVLTGMEKGGQLTTTTEVENRFGDNDLT
 GFTKHSKLLIGSSRGATTANTYAAANLQPVLTGMEKGGQLTTTTEVENRFGDNDLT
 GFTKHSKLLIGSSRGATTANTYAAANLQPVLTGMEKGGQLTTTTEVENRFGDNDLT
 GFTKHSKLLIGSSRGATTANTYAAANLQPVLTGMEKGGQLTTTTEVENRFGDNDLT
 LPESEAFKGRGVSACATCDGFFRNQKQVAVIGGGNTAVEALYLSNIASEVHLIHRDGF
 RAKKILIKRLMDKVENGNILHTNRTLEEVTGQDQWVTGVRRLRUTQNSDNIESLDVAGLF
 VAIGHSPNTAIFBQGLELENGYIKVQSGIHGNATQTSIFGVFAAGDMDHRYRQAITSAG
 TGCMAALDAERYLGLADAK

Figure 21B

>sp|P08860|TRXB_BACSU Thioredoxin reductase (EC 1.6.4.5) (TRXR) (General stress protein 35) (GSP35) - Bacillus subtilis.
 SEEKIYDVIIIGAGPAGMTAAVYTSKANLSTIMIERGTPGGQMANTEVDVNYGPFESILG
 PELSNNMFEHAKKFGARYAYGDIKEVIDGKEYKVKAGSKEYKARAVLIAAGAEYKKIGV
 PGBSELGGRGVSICAVCDGAFPRGRLVVGSGDSABEGVILTRFASKVTIVHRRDKLR
 AGSILQARAFQNKVDFLWAKYVKELEHBNKQVGVVILVDTVTGESESKTDEVEIILGM
 LPLSKFPEHLGILTNEBGYIETNDRMETKVEGIFAGDIRKSLRQIVTATGDSANQSV
 QHYVEELQETLTKLK

Figure 21C

>sp|P46843|TRXB_MYCLE Bifunctional thioredoxin reductase/thioredoxin [Includes: Thioredoxin reductase (EC 1.6.4.5) (TRXR); Thioredoxin] - Mycobacterium leprae.
 NHTTFSRHETHEVIVIGSPGYTAALYARQLTFLVFGTSFGCALMTTTEVENTPG
 FRNGITGPELMDNREQALRFGAELRTEDVESVLRGPIKSVYTAEGQYQANVILANG
 TSVRYLQIPGEOELLGRGVSACATCDGSSFFRGQDIIVIGGDSAMEALFLTRFASVTL
 VHRDEFRASKMLGRARNNDKIKFTNHTVAVNGYITVGLRLANTTGEETILVITG
 VFAVIGHEPRSSINSDVDVIDPDGVLVKGRTTSTMDGVFAAGDLVDRTYRQAITSAG
 GCAAIIDAERWLAEIAGSKANETTBETGHDVDSFTUTLWSTAMTDARNAAGVTIEVTASFF
 ADVLSNKEVIVDWTATWCGPCRCVAVPLEIASQKQGLTVAKLVDTNFMHAREFQVV
 SIPTMILFDGGQPVKRIYAGKGAALLDLSDVVPNLN

Figure 21D

>sp|P51978|TRXB_NEUCR Thioredoxin reductase (EC 1.6.4.5) - Neurospora crassa.
 MHKSVVIGSGPAHATAIYLARAEKLPVLYEGFMNGIAAGGQLTTTTRIENFPGFPDG
 IMGELMDKMKQSRERFGTQISETVAKVDSARFFVATENSPEEYHTADSIIILATGAS
 ARRLMLPGEETVWQKGISACAVCDGAVPIFRNKEILVVGSGDSAAEEAMILTYSKSHVT
 LVKDKLRASINARILNNEKVTVEFNTVGVGVKGDQKGLMSLHVKDVTTGKEZTLEA
 NGLEYALGHDPALVKGQLETDADGVTYVTKPGTTLTSVEGVFAAGDVQDKRYRQAITS
 GTGCMAALDAEKFLSHHEETPAHHRDTSVQGNL

Figure 21E

>sp|P29509|TRB1_YEAST Thioredoxin reductase 1 (EC 1.6.4.5) - Saccharomyces cerevisiae (Baker's yeast).
 VHNKVTIIIGSGPAHATAIYLARAEKLPVLYEGFMNGIAAGGQLTTTTRIENFPGFPDG
 LTGSSILMDRMREQSTKFGTEIITETVSKVDLSKPKFLWTEFNEDAEPTTDAIILATGA
 SAKRMHLPGEETVWQKGISACAVCDGAVPIFRNKEILVVGSGDSAAEEAMILTYSKSHVT
 MLVRKDLRSTIMOKRAEKNEKIELVNTVALEAKGDGKLLNALIKNIKNESTDLV
 SGLFIAIGHTPAIKIYAGQVUTDEAGYIKTVPGSSLSVPGFFAAGDVQDKRYRQAITS
 GSGCMAALDAEKFLSHHEETPAHHRDTSVQGNL

Figure 21F

>sp|P38816|TRB2_YEAST Thioredoxin reductase 2, mitochondrial precursor (EC 1.6.4.5) - Saccharomyces cerevisiae (Baker's yeast).
 MTKHIVSPFRTNVIGISKSVLSRMTHHKVTIIIGSGPAHATAIYLARAEKLPVLYEGFMN
 NGIAGGQLTTTTRIENFPGFPESILSGSLMDRMKQSAFGTITITETVSKVDLSKPKFLW
 RTEFNEDAEPTTDAIILATGASAKRMHLPGEETVWQKGISACAVCDGAVPIFRNKEIL
 AVIGGDSACEAEFLTKYASKVYILVRKDLFRASVIMORRIEKNENIIVLNTVALRAK
 GDGKLLNMLRIKNTKSNVENDLVNGLFYAIGHSFATDVLKGVDERETGYIKTVPGSSLS
 TSVPGFFAAGDVQDSRYQAVTSAGSGCIAALDAERYLSAQE

Figure 21G

>sp|Q39243|TRB1_ARATH Thioredoxin reductase 1 (EC 1.6.4.5) (NADPH-dependent thioredoxin reductase 1) (NTR 1) - Arabidopsis thaliana (Mouse-ear cress).

WO 02/090300

PCT/US02/14358

MNGLETHNTRLCIVGSGPAHTAAIYAPAEKPLLFEGWANDIAFGGQLTTTTDVENF
PGFPESILGVELTDKFRKQSERFGTTIFETVTRVDFSSKPFKLTDSKAILADAVILAT
GAVAKRLSEVGSSEASGCFWNRGISACAVCDGARPIFRKPLAVIGGDSAMELANELTK
YGSKVYIIHRDAFRASKIMQORALSNPKIDVINSSVVAVGDSERDVLGGLAKVWVT
GDVSDLKVSLGFAIGREEATKFLDGGVELSDGYVVKFGTQTSTVPGVFAAGDVQDKK
YRQAITAAGTGCMAALDAEHLQETIGSQGKSD

Figure 21H

>sp|Q59242|TRB2_ARATH Thioredoxin reductase 2 (EC 1.6.4.5) (NADPH-dependent
thioredoxin reductase 2) (NTR 2) - Arabidopsis thaliana (Mouse-ear cress).
MCWTSMGQSRFTIKSLFSTAGGFLGSLSNFPFLATAPSSSSSSSSAAAADVMEHKYK
VCIVGSGGPAHTAAIYASPAELKPLLFEGWANDIAFGGQLTTTTDVENFPGFPGCILGI
DIVEKFRKQSERFGTTIFETVTRVDFSSKPFKLTDSRTVLADSVIISTGAVAKRLSP
GSGGNGCGFWNRGISACAVCDGARPIFRNKLPLVIGGDSAMREANFI/KYGSKVYLIHR
RDTFRASLIMQORALSNPKIEVINSAVVSAYGDENGRVLGGLKVRNVITGDSDLKVS
LFFAIGHRPATKTLGQLLEDDGYVVKFGTQTSTVPGVFAAGDVQDKKYRQAITAAGT
GCMAALDAEHLQETIGSQGKSD

Figure 21I

>sp|Q16881|TRXB_HUMAN Thioredoxin reductase (EC 1.6.4.5) - Homo sapiens (Human).
MNGFEDLEKSYDYDLIIIGSGGGLAAAEQAQYKKVMVLDVFTPTPLGTRWGLGGTCV
NVGCIPEKLMHQAALIGALQDSRNYGWRVEETVEHMDRMTEAVQNHIGSLNMGYRVAL
REKVVYENAGQPTGPHRLATNWKGEKITYASSLIATGERPRYLGI PGCEGYCIS
DDLFSIPYPCPKPLVVGASYVALEKAGELAGIGLVTVWVRSILRGFDQDMANKIGEHM
EEHGKIFTRQFVPIKVQETIAGTFGRLRVVAQSTNSEETIEGEYNTVMLAIGRDACTRI
GLETVGVKINEKTKIPVTEEBQTNVPIYVAGDILEDKVELTPVAIQGRLLAQRLYAG
STVKCDYENVFTVFTFLEYGACGLSEKAVEKFGENIEVYHSYFPLEWTIPERDNNK
CTAKIICNTKDERVVGFFVLGFENAGEVTQGFRAALKCGLKKQLDSTIGTHPVCHEVPT
TLSVTRSGSILQAGC

PCT/US02/14358

10

111/113

WO 02/090300

PCT/US02/14358

Figure 22 continued

```

      220      230      240      250      260      270
      |       |       |       |       |       |
MG..VTGVRLRDTQNSDNIES.L...DVRGLFVAIGHSPNTAIFEG.QL.EL.E.NGVYKVSQGIH...
GK..VGNVTLVDIVTQGESE-.F...XTDGVETIYICMLPLSKPFENIGI.TN.E.EGVYET-----
TT..VTGLRLRNTTQGESE-TT.L...VVTGVFVAIGHSPRSSLVSD.VV.DI.DpDGVVVK-----
KG..LMSHLAVKDVITGKEET.L...FANGLFYAIGHDPATLVKG.QL.ET.DaDGVVVKPG-----
KL..LAMLRIKNTKKNNEED-.L...PVSGLFYAIGHDPATKIVAG.QV.DTDE.AGYIKTVFEG-----
KL..LAMLRIKNTKKNNEED-.L...PVSGLFYAIGHDPATKIVAG.QV.DTDE.AGYIKTVFEG-----
DV..LCGLKVGWVVICD-VSD.L...KVSGLFFAIGHSPATKIVAG.QV.DL.DaDGVVVKPG-----
NGVILGGLKVKVVICD-VSD.L...KVSGLFFAIGHSPATKIVAG.QL.ELDE.DGVVVKPG-----
AG..TPG-RLRVVAGSTNSSE.LieqEYNFVMLAIGRDACR----.KI.GL.E.TVGVKINEKTEkipv

      280      290      300      310      320
      |       |       |       |       |
GNATQTSIPGVFAAGDVMDHIYRQALTSAGTGCMAALDAERYLOGIADAK.....
NDRMETKVBGIFAAGDIREKSLRQIVTATGDSIAAGSVQHYVEILQRTLkklk.....
GRTTSTSMGDFVFAAGDIVDRTYRQAITAAGSGCAAIADARRNLAEHAGSKanetteetgvdstdtdtdws
--TTLTSVEGVFAAGDVQDKRYRQALTSAGTGCMAALDAKFLSEHEETPaehrdtsavqgnl.....
--SELTSVPGFFAAGDVQDKRYRQALTSAGSGCAALDAKFLSEHEETPaehrdtsavqgnl.....
--SELTSVPGFFAAGDVQDKRYRQALTSAGSGCAALDAERYL----SAQSE.....
--TTSVPGVFAAGDVQDKRYRQALTSAGTGCMAALDAERYLQELIGSQGksd.....
--TTSVPGVFAAGDVQDKRYRQALTSAGTGCMAALDAERYLQELIGSQGksd.....
TDESQTNVPYIYAIGDILLEKVELTFAIQAG---RLAQRIYAG-STVKedyenvpttvtitpleygaag

.....
tamtdaknagvtievtasffadlssnkpvlvdfwatwcpckmwavpveeiaseqrnqltvakldvdt
.....
.....
lseekavekfgeenievyhsyfwplewtipsrdnnkcyakilcntkdnervvgfhvlgpnagevtqgfaa

.....
npemerefqvvsipnlieggqgvitvgaakkaalirdsdvvnln.
.....
.....
aikogltkklqldstiglhpvcaewftlsvtkregasillgagc.....

```

WO 02/090300

PCT/US02/14358

Figure 21J

trxB from *Methanococcus jannaschii* (gil592167):

MIHDTTIIIGAGPGGLTAGIYAMRGKLNALCIEKENAGGRIAEAGIVENYFGFEEIRGYELAEKP
KNHAEEKFLPIIYDEVIKIETKRRFPKVIITKNSEYLTKTIVLATGTKPKKLGLNEDKFTGRGIS
YCMCDAFFYLNGEVIVIGRDTPAINSAINLKDIAKKVIVITDKSELKAARSINLDDKLEANNV
EIIYNAPLEIVGSEBRAEGVKISVNGKEELIKADGIFISLGHVFNTEFLKDSGIELDKKGFIKT
DENCRINIDGIYAVGDVRGGVMQVAKAVGDCVAMANI IKYLQKL

Figure 21K

trxB from *Amhaeoglobus fulgidus* (gil2649006):

MYDVAIIIGGPAGLPAALYSARYGLKIVFFETVDPVSQLSLAAKTENYFGFEGSMELLEEMKE
QAVKAGAEWKLEKVERVRNGETFTVIAEGGEYEAKEIIVATGCKHKEAGIGBSAFIGRGVSY
CATCDGNFPRGKKVIVYSGKRAIEDAILYLDIGCEVTIVSRTPSPRAEKALVBEVEKNGIPVH
YSTTIRKIIGSGKVRKVVAYNREKKKEETIRADGIFVAIGMRPATDVVAELGVERDSMGYIKVD
KBQRTNVEGVPAAGDCCDNPLKQVVTACGDAVAAYSAKYKITS.

【国際公開パンフレット（コレクトバージョン）】

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

CORRECTED VERSION

(19) World Intellectual Property
Organization
International Bureau



(43) International Publication Date
14 November 2002 (14.11.2002)

PCT

(10) International Publication Number
WO 2002/090300 A2

- (51) International Patent Classification⁷: **C07C**
- (21) International Application Number: PCT/US2002/014358
- (22) International Filing Date: 6 May 2002 (06.05.2002)
- (25) Filing Language: English
- (26) Publication Language: English
- (30) Priority Data:
60/289,029 4 May 2001 (04.05.2001) US
60/370,609 5 April 2002 (05.04.2002) US
60/376,682 29 April 2002 (29.04.2002) US
- (71) Applicants: XENCOR [US/US]; 111 West Lemon Avenue, Monrovia, CA 91016 (US); SYNGENTA PARTICIPATIONS AG [CH/CH]; Schwarzwaldallee 215, CH-4058 Basel (CH).
- (72) Inventors: BRIGGS, Steven, P.; 2454 Mango Way, Del Mar, CA 92014 (US); DALMIA, Bipin, K.; 7353 Mannix Court, San Diego, CA 92129 (US); DEL VAL, Gregory; 5727 Erlangen Street, San Diego, CA 92122 (US); DESJARLAIS, John, R.; 2096 Crary Street, Pasadena, CA 91104 (US); HEIFETZ, Peter; 10805 Birch Bluff Avenue, San Diego, 92131 (US); LUGINBUHL, Peter; 13567 Jadestone Way, San Diego, CA 92130 (US); MUCHHAL, Umesh; 200 North Grand Avenue, #248, West Covina, CA 91291 (US).
- (74) Agents: TRECARTIN, Richard, F. et al.; Dorsey & Whitney LLP, Suite 3400, 4 Embarcadero Center, San Francisco, CA 94111-4187 (US).
- (81) Designated States (national): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, UZ, VN, YU, ZA, ZM, ZW.
- (84) Designated States (regional): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LI, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).
- Published:
— without international search report and to be republished upon receipt of that report
- (48) Date of publication of this corrected version: 22 April 2004
- (15) Information about Correction:
see PCT Gazette No. 17/2004 of 22 April 2004, Section II
- For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.



WO 2002/090300 A2

(54) Title: NUCLEIC ACIDS AND PROTEINS WITH THIOREDOXIN REDUCTASE ACTIVITY

(57) Abstract: The present invention relates to the use of a variety of methods for generating functional thioredoxin reductase variants in which at least one physical, chemical or biological property of the variant is altered in a specific and desired manner when compared to the wild-type protein.

フロントページの続き

(51)Int.Cl. ⁷	F I	テーマコード (参考)
C 1 2 P 7/64	C 1 2 P 7/64	4 B 0 6 4
G 0 6 F 17/30	G 0 6 F 17/30	4 H 0 4 5
// A 2 3 L 1/30	A 2 3 L 1/30	5 B 0 7 5
		B

(81)指定国 AP(GH,GM,KE,LS,MW,MZ,SD,SL,SZ,TZ,UG,ZM,ZW),EA(AM,AZ,BY,KG,KZ,MD,RU,TJ,TM),EP(AT,BE,CH,CY,DE,DK,ES,FI,FR,GB,GR,IE,IT,LU,MC,NL,PT,SE,TR),OA(BF,BJ,CF,CG,CI,CM,GA,GN,GQ,GW,ML,MR,NE,SN,TD,TG),AE,AG,AL,AM,AT,AU,AZ,BA,BB,BG,BR,BY,BZ,CA,CH,CN,CO,CR,CU,CZ,DE,DK,DM,DZ,EC,EE,ES,FI,GB,GD,GE,GH,GM,HR,HU,ID,IL,IN,IS,JP,KE,KG,KP,KR,KZ,LC,LK,LR,LS,LT,LU,LV,MA,MD,MG,MK,MN,MW,MX,MZ,NO,NZ,OM,PH,PL,PT,RO,RU,SD,SE,SG,SI,SK,SL,TJ,TM,TR,TT,TZ,UA,UG,UZ,VN,YU,ZA,ZM,ZW

(74)代理人 100067035

弁理士 岩崎 光隆

(74)代理人 100064610

弁理士 中嶋 正二

(74)代理人 100072730

弁理士 小島 一晃

(72)発明者 スティーブン・ピー・ブリッグズ

アメリカ合衆国 9 2 0 1 4 カリフォルニア州デル・マー、マンゴー・ウェイ 2 4 5 4 番

(72)発明者 ビピン・ケイ・ダルミア

アメリカ合衆国 9 2 1 2 9 カリフォルニア州サンディエゴ、マニックス・コート 7 3 5 3 番

(72)発明者 グレゴリー・デル・バル

アメリカ合衆国 9 2 1 2 2 カリフォルニア州サンディエゴ、アーランジェン・ストリート 5 7 2 7 番

(72)発明者 ジョン・アール・デスジャーレイス

アメリカ合衆国 9 1 1 0 4 カリフォルニア州パサディナ、クラリー・ストリート 2 0 9 6 番

(72)発明者 ピーター・ハイフェッツ

アメリカ合衆国 9 2 1 3 1 カリフォルニア州サンディエゴ、バーチ・ブラフ・アベニュー 1 0 8 0 5 番

(72)発明者 ピーター・ルジンブール

アメリカ合衆国 9 2 1 3 0 カリフォルニア州サンディエゴ、ジェイドストーン・ウェイ 1 3 5 6 7 番

(72)発明者 ウメシュ・マッチハル

アメリカ合衆国 9 1 2 9 1 カリフォルニア州ウエスト・コビナ、ノース・グランド・アベニュー 2 0 0 番、ナンバー 2 4 8

Fターム(参考) 2B030 AA02 AD08 CA14 CB02

2B150 AC37 CJ07 DD31

4B018 LB01 LB02 LB04 LB05 LB06 LB07 ME07 ME14

4B024 AA05 BA08 BA80 CA04 CA07 DA01 DA02 DA05 DA11 EA01

EA02 EA03 EA04 FA01 GA11 HA01 HA03

4B050 CC03 CC05 DD02 DD04 DD11 DD13 LL02

4B064 AD85 AG01 CA01 CA19 CC24 DA10

4H045 AA30 BA10 BA41 CA10 CA11 CA30 CA40 DA89 EA01 FA72

FA74

5B075 UU18