(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2004-510440
(P2004-510440A)
(43)公表日　平成16年4月8日(2004.4.8)

(51)Int.Cl.[7]	F I		テーマコード（参考）
C12N 15/09	C12N 15/00	ZNAA	4B024
A61K 48/00	A61K 48/00		4B064
A61P 1/04	A61P 1/04		4C084
A61P 3/10	A61P 3/10		4C085
A61P 5/14	A61P 5/14		4C087

審査請求 未請求　予備審査請求 有　（全 134 頁）　最終頁に続く

(21)出願番号　特願2002-532657(P2002-532657)
(86)(22)出願日　平成13年10月4日(2001.10.4)
(85)翻訳文提出日　平成15年4月2日(2003.4.2)
(86)国際出願番号　PCT/US2001/031451
(87)国際公開番号　WO2002/029088
(87)国際公開日　平成14年4月11日(2002.4.11)
(31)優先権主張番号　60/237,885
(32)優先日　平成12年10月4日(2000.10.4)
(33)優先権主張国　米国(US)

(71)出願人　500429103
ザ・トラスティーズ・オブ・ザ・ユニバーシティ・オブ・ペンシルバニア
アメリカ合衆国ペンシルバニア州１９１０４－６２８３，フィラデルフィア，チェスナット・ストリート　３１６０，スウィート　２００
(74)代理人　100104411
弁理士　矢口　太郎
(74)代理人　100104215
弁理士　大森　純一
(74)代理人　100099656
弁理士　山口　康明

最終頁に続く

(54)【発明の名称】高発現可能遺伝子

(57)【要約】
【解決手段】本発明は組換え発現系において蛋白質を産生する方法に関するものであり、発現系において、異種ＤＮＡ配列から転写されるｍＲＮＡを翻訳する過程から成り立つ。本方法は野生型異種ＤＮＡ配列から転写されるｍＲＮＡの二次構造を予測する工程を含む。この修飾型異種ＤＮＡ配列から転写されるｍＲＮＡの二次構造は、野生型異種ＤＮＡ配列から転写されるｍＲＮＡの二次構造に比べて、増大した自由エネルギーを有する。さらに、本方法は修飾型異種ＤＮＡ配列を用いて、組換え発現系において蛋白質を産生する工程を含む。本発明はまた、修飾型コード配列を含む核酸分子からなる注射可能な医薬成分を提供する。本発明はまたコード配列も提供する。

【特許請求の範囲】
【請求項１】
組換え発現系における蛋白質産生に関する方法であって、発現系における異種ＤＮＡ配列から転写されるｍＲＮＡの翻訳によって成り立ち、この方法は、
ａ）野生型異種ＤＮＡ配列から転写されるｍＲＮＡの二次構造を予測する工程と、
ｂ）野生型異種ＤＮＡから転写されるｍＲＮＡの二次構造に比較して、修飾型異種ＤＮＡ配列から転写されるｍＲＮＡの二次構造が増大した自由エネルギーを有するように、野生型異種ＤＮＡ配列を修飾して、修飾型異種ＤＮＡ配列を生産する工程と、
ｃ）蛋白質生産の為に、組換え発現系において修飾型異種ＤＮＡ配列を用いる工程とを有する。
【請求項２】
請求項１に記載の方法における組換え発現系は、無細胞試験管内転写並びに翻訳の系、試験管内細胞発現系、ＤＮＡ直接注入に用いられるＤＮＡ構築物、個体にＤＮＡを導入する為に用いる組換えベクターからなる群より選択される。
【請求項３】
請求項１に記載の方法において、野生型異種ＤＮＡ配列から転写されるｍＲＮＡの二次構造は、コンピューターとコンピュータープログラムを用いて予測される。
【請求項４】
請求項１に記載の方法において、野生型異種ＤＮＡ配列から転写されるｍＲＮＡの二次構造は、コード配列のアデニンとチミンの含有量を増加することによって修飾される。
【請求項５】
請求項４に記載の方法において、野生型異種ＤＮＡ配列から転写されるｍＲＮＡの二次構造は、コード配列の５’末端においてコード配列のアデニンとチミンの含有量を増加することにより、そこから転写されるｍＲＮＡがより多くのアデニンとウラシルとを含有するように修飾される。
【請求項６】
請求項５に記載の方法において、野生型異種ＤＮＡ配列から転写されるｍＲＮＡの二次構造は、開始コドンより２００ヌクレオチド以内のコード配列の５’末端においてコード配列のアデニンとチミンの含有量を増加することにより、そこから転写されるｍＲＮＡがより多くのアデニンとウラシルとを含有するように修飾される。
【請求項７】
請求項６に記載の方法において、野生型異種ＤＮＡ配列から転写されるｍＲＮＡの二次構造は、開始コドンより１５０ヌクレオチド以内のコード配列の５’末端においてコード配列のアデニンとチミンの含有量を増加することにより、そこから転写されるｍＲＮＡがより多くのアデニンとウラシルとを含有するように修飾される。
【請求項８】
請求項６に記載の方法において、野生型異種ＤＮＡ配列から転写されるｍＲＮＡの二次構造は、開始コドンより１００ヌクレオチド以内のコード配列の５’末端においてコード配列のアデニンとチミンの含有量を増加することにより、そこから転写されるｍＲＮＡがより多くのアデニンとウラシルとを含有するように修飾される。
【請求項９】
調節エレメントと連結して用いることが可能で、蛋白質をコードする修飾型コード配列を含む核酸分子を有する注射可能な医薬品成分であり、前記の修飾型コード配列は、野生型コード配列のアデニンとチミンもしくはアデニンとウラシルの含有量に比較して、より高いアデニンとチミンもしくはアデニンとウラシルを含有量を有しており、さらに医薬的に受容可能な担体を有する注射可能な医薬品成分。
【請求項１０】
請求項９に記載の注射可能な医薬品成分において、前記の修飾型コード配列は、野生型核酸配列のアデニンとチミンもしくはアデニンとウラシルの含有量と比較して、最初の２００塩基においてより高いアデニンとチミンもしくはアデニンとウラシルの含有量を有する

。
【請求項１１】
請求項９に記載の注射可能な医薬品成分において、前記の修飾型コード配列は、野生型核酸配列のアデニンとチミンもしくはアデニンとウラシルの含有量と比較して、最初の１５０塩基においてより高いアデニンとチミンもしくはアデニンとウラシルの含有量を有する。
【請求項１２】
請求項９に記載の注射可能な医薬品成分において、前記の修飾型コード配列は、野生型核酸配列のアデニンとチミンもしくはアデニンとウラシルの含有量と比較して、最初の１００塩基においてより高いアデニンとチミンもしくはアデニンとウラシルの含有量を有する。
【請求項１３】
請求項９に記載の注射可能な医薬品成分において、前述の修飾型コード配列は、野生型核酸配列のアデニンとチミンもしくはアデニンとウラシルの含有量と比較して、２００塩基長までの少なくともひとつの領域において、より高いアデニンとチミンもしくはアデニンとウラシルの含有量を有する。
【請求項１４】
請求項９に記載の注射可能な医薬品成分において、前記の修飾型コード配列は、野生型核酸配列のアデニンとチミンもしくはアデニンとウラシルの含有量と比較して、１５０塩基長までの少なくともひとつの領域において、より高いアデニンとチミンもしくはアデニンとウラシルの含有量を有する。
【請求項１５】
請求項９に記載の注射可能な医薬品成分において、前記の修飾型コード配列は、野生型核酸配列のアデニンとチミンもしくはアデニンとウラシルの含有量と比較して、１００塩基長までの少なくともひとつの領域において、より高いアデニンとチミンもしくはアデニンとウラシルの含有量を有する。
【請求項１６】
請求項９に記載の注射可能な医薬品成分において、前記の修飾型コード配列は免疫原をコードしている。
【請求項１７】
請求項１６に記載の注射可能な医薬品成分において、前記免疫原は病原体由来の蛋白質またはそこからの免疫原性断片である。
【請求項１８】
請求項１６に記載の注射可能な医薬品成分において、前記免疫原は病原体由来の蛋白質またはそこからの免疫原性断片を含む融合蛋白質である。
【請求項１９】
請求項１６に記載の注射可能な医薬品成分において、前記免疫原は癌抗原またはそこからの免疫原性断片である。
【請求項２０】
請求項１６に記載の注射可能な医薬品成分において、前記免疫原は癌抗原またはそこからの免疫原性断片を含む融合蛋白質である。
【請求項２１】
請求項１６に記載の注射可能な医薬品成分において、前記免疫原は自己免疫性疾患関連蛋白質またはそこからの免疫原性断片である。
【請求項２２】
請求項１６に記載の注射可能な医薬品成分において、前記免疫原は自己免疫疾患関連蛋白質またはそこからの免疫原性断片を含む融合蛋白質である。
【請求項２３】
請求項９に記載の注射可能な医薬品成分において、前記の修飾型コード配列は非免疫原性治療蛋白質である。

【請求項２４】
請求項２３に記載の注射可能な医薬品成分において、前記非免疫原性治療蛋白質は、サイトカイン、成長因子、血液産物、酵素から成るグループから選ばれたものである。
【請求項２５】
請求項９に記載の注射可能な医薬品成分において、前記の修飾型コード配列は分散修飾を含む。
【請求項２６】
請求項２５に記載の注射可能な医薬品成分において、前記分散修飾は、２００塩基長までの少なくとも２つの修飾コード配列であり、野生型コード配列の領域と交互に配置される。
【請求項２７】
請求項２５に記載の注射可能な医薬品成分において、前記分散修飾は、１５０塩基長までの少なくとも２つの修飾コード配列であり、野生型コード配列の領域と交互に配置される。
【請求項２８】
請求項２５に記載の注射可能な医薬品成分において、前記分散修飾は、１００塩基長までの少なくとも２つの修飾コード配列であり、野生型コード配列の領域と交互に配置される。
【請求項２９】
請求項９に記載の注射可能な医薬品成分において、前記修飾型コード配列は、野生型核酸配列のアデニンとチミンまたはアデニンとウラシルの含有量と比較して、最後の２００塩基においてより高いアデニンとチミンまたはアデニンとウラシルの含有量を有する。
【請求項３０】
請求項９に記載の注射可能な医薬品成分において、前記の修飾型コード配列は、野生型核酸配列のアデニンとチミンまたはアデニンとウラシルの含有量と比較して、最後の１５０塩基においてより高いアデニンとチミンまたはアデニンとウラシルの含有量を有する。
【請求項３１】
請求項９に記載の注射可能な医薬品成分において、前記の修飾型コード配列は、野生型核酸配列のアデニンとチミンまたはアデニンとウラシルの含有量と比較して、最後の１００塩基においてより高いアデニンとチミンまたはアデニンとウラシルの含有量を有する。
【請求項３２】
調節エレメントと連結して用いることが可能で、蛋白質をコードする修飾型コード配列を含む核酸分子を有する組換えウイルスベクターであって、前記の修飾型コード配列は、野生型コード配列のアデニンとチミンもしくはアデニンとウラシルの含有量に比較して、より高いアデニンとチミンもしくはアデニンとウラシルを含有量を有する。
【請求項３３】
請求項３２に記載の組換えウイルスベクターにおいて、前記修飾型コード配列は、野生型核酸配列のアデニンとチミンまたはアデニンとウラシルの含有量と比較して、最初の２００塩基においてより高いアデニンとチミンまたはアデニンとウラシルの含有量を有する。
【請求項３４】
請求項３２に記載の組換えウイルスベクターにおいて、前記修飾型コード配列は、野生型核酸配列のアデニンとチミンまたはアデニンとウラシルの含有量と比較して、最初の１５０塩基においてより高いアデニンとチミンまたはアデニンとウラシルの含有量を有する。
【請求項３５】
請求項３２に記載の組換えウイルスベクターにおいて、前記修飾型コード配列は、野生型核酸配列のアデニンとチミンまたはアデニンとウラシルの含有量と比較して、最初の１００塩基においてより高いアデニンとチミンまたはアデニンとウラシルの含有量を有する。
【請求項３６】
請求項３２に記載の組換えウイルスベクターにおいて、前記修飾型コード配列は、野生型核酸配列のアデニンとチミンまたはアデニンとウラシルの含有量と比較して、２００塩基

長までの少なくともひとつの領域においてより高いアデニンとチミンまたはアデニンとウラシルの含有量を有する。
【請求項３７】
請求項３２に記載の組換えウイルスベクターにおいて、前記修飾型コード配列は、野生型核酸配列のアデニンとチミンまたはアデニンとウラシルの含有量と比較して、１５０塩基長までの少なくともひとつの領域においてより高いアデニンとチミンまたはアデニンとウラシルの含有量を有する。
【請求項３８】
請求項３２に記載の組換えウイルスベクターにおいて、前記修飾型コード配列は、野生型核酸配列のアデニンとチミンまたはアデニンとウラシルの含有量と比較して、１００塩基長までの少なくともひとつの領域においてより高いアデニンとチミンまたはアデニンとウラシルの含有量を有する。
【請求項３９】
請求項３２に記載の組換えウイルスベクターにおいて、前記修飾型コード配列は免疫原をコードしている。
【請求項４０】
請求項３９に記載の組換えウイルスベクターにおいて、前記免疫原は病原体由来蛋白質またはそこからの免疫原性断片である。
【請求項４１】
請求項３９に記載の組換えウイルスベクターにおいて、前記免疫原は病原体由来蛋白質またはそこからの免疫原性断片を含む融合蛋白質である。
【請求項４２】
請求項３９に記載の組換えウイルスベクターにおいて、前記免疫原は癌抗原またはそこからの免疫原性断片である。
【請求項４３】
請求項３９に記載の組換えウイルスベクターにおいて、前記免疫原は癌抗原またはそこからの免疫原性断片を含む融合蛋白質である。
【請求項４４】
請求項３９に記載の組換えウイルスベクターにおいて、前記免疫原は自己免疫疾患関連蛋白質またはそこからの免疫原性断片である。
【請求項４５】
請求項３９に記載の組換えウイルスベクターにおいて、前記免疫原は自己免疫疾患関連蛋白質またはそこからの免疫原性断片を含む融合蛋白質である。
【請求項４６】
請求項３９に記載の組換えウイルスベクターにおいて、前記修飾型コード配列は非免疫原性治療蛋白質をコードしている。
【請求項４７】
請求項４６に記載の組換えウイルスベクターにおいて、前記非免疫原性治療法蛋白質はサイトカイン、成長因子、血液産物、酵素から成るグループから選ばれたものである。
【請求項４８】
請求項３２に記載の組換えウイルスベクターにおいて、前記修飾型コード配列は分散修飾を含む。
【請求項４９】
請求項４８に記載の組換えウイルスベクターにおいて、前記分散修飾は、野生型コード配列の領域と交互に配列される、２００塩基長の少なくとも２つの修飾コード配列である。
【請求項５０】
請求項４８に記載の組換えウイルスベクターにおいて、前記分散修飾は、野生型コード配列の領域と交互に配列される、１５０塩基長の少なくとも２つの修飾コード配列である。
【請求項５１】
請求項４８に記載の組換えウイルスベクターにおいて、前記分散修飾は、野生型コード配

列の領域と交互に配列される、１００塩基長の少なくとも２つの修飾コード配列である。
【請求項５２】
請求項３２に記載の組換えウイルスベクターにおいて、前記修飾型コード配列は、野生型核酸配列のアデニンとチミンまたはアデニンとウラシルの含有量と比較して、最後の２００塩基においてより高いアデニンとチミンまたはアデニンとウラシルの含有量を有する。
【請求項５３】
請求項３２に記載の組換えウイルスベクターにおいて、前記修飾型コード配列は、野生型核酸配列のアデニンとチミンまたはアデニンとウラシルの含有量と比較して、最後の１５０塩基においてより高いアデニンとチミンまたはアデニンとウラシルの含有量を有する。
【請求項５４】
修飾型コード配列は、野生型核酸配列のアデニンとチミンまたはアデニンとウラシルの含有量と比較して、最後の１００塩基においてより高いアデニンとチミンまたはアデニンとウラシルの含有量を有する。
【発明の詳細な説明】
【０００１】
本出願は、３５ Ｕ．Ｓ．Ｃ．§１１９（ｅ）に基づき、引用により本明細書の一部に組み込まれている２０００年１０月４日付け出願の米国仮出願番号６０／２３７，８８５号の利益を主張するものである。
【０００２】
【発明の属する技術分野】
この発明は、一般的に遺伝子発現、遺伝子治療、遺伝免疫の分野に関するものである。
【０００３】
【発明の背景】
遺伝子産物としての蛋白質の発現は、遺伝子コピー数、ゲノム中の遺伝子組み込み部位もしくは遺伝子の位置、転写因子、ｍＲＮＡの安定性、翻訳効率を含む多くの要因の影響を受ける。例えば、ヒト免疫不全ウイルス－１（ＨＩＶ－１）の構造遺伝子ｇａｇ、ｐｏｌ、ｅｎｖの発現は、転写後のレベルにおいてＲｅｖ／Ｒｅｖ反応エレメント（ＲＲＥ）に依存している。このＲｅｖへの依存性は遺伝子発現に対する制限因子となる。加えて、ＨＩＶのＲＮＡ転写産物の様々な領域において形成される、非常に安定なＲＮＡの二次構造は、リボゾームの動きを阻害、あるいは干渉し、効果的に翻訳を制限する。遺伝子転写産物の安定なＲＮＡ二次構造の形成は一般的な現象であり、それによって、広範な種類の遺伝子について、遺伝子産物すなわち蛋白質への翻訳量が制限される。
【０００４】
ここに参考文献として含められる、Ｇｅｎｅ，１９９；２９３－３０１（１９９７年）において、キムらは哺乳類の細胞におけるヒトエリスロポエチン（ＥＰＯ）の発現を至適化する為に、成熟ＥＰＯ蛋白質のリーダー配列と最初の６アミノ酸をコードするコドンを酵母細胞で最も汎用されているコドンに変更し、成熟ＥＰＯ蛋白質のそれ以外の部分をコードするコドンをヒトで最も汎用されているコドンに変更した。
【０００５】
ここに参考文献として含められる、米国特許番号５，９７２，５９６と５，９６５，７２６（パブラキス（Ｐａｖｌａｋｉｓ）ら）は、ｍＲＮＡのコード領域内にみられる阻害性あるいは不安定性の配列（ＩＮＳ：ｍＲＮＡを不安定、あるいは十分に利用できなくする、または翻訳できなくする配列）の局在を決定する方法と、ヌクレオチドの群置換により、ＩＮＳを取り除く目的で、ｍＲＮＡをコードしている遺伝子を修飾する方法を述べている。
【０００６】
蛋白質を発現する為の新しい方法が必要であり、治療目的ならびに免疫原性導入遺伝子の蛋白質発現レベルを増加する為の新しい方法が必要である。任意の蛋白質性遺伝子産物の翻訳収量を増加する為の方法が必要である。ＲＮＡ転写産物における、ＲＮＡの二次構造によってもたらされる、任意の遺伝子の蛋白質発現レベルの限界を克服する方法が必要で

ある。本発明は、これらならびに他の必要性に応えようとするものである。
【０００７】
【発明の概要】
本発明は組換え発現系において蛋白質を産生する方法に関するものであり、発現系において、異種ＤＮＡ配列から転写されるｍＲＮＡを翻訳する過程から成り立つ。本方法は野生型異種ＤＮＡ配列から転写されるｍＲＮＡの二次構造を予測する工程、野生型異種ＤＮＡ配列を修飾することにより、修飾型異種ＤＮＡ配列を生産する工程を含む。この修飾型異種ＤＮＡ配列から転写されるｍＲＮＡの二次構造は、野生型異種ＤＮＡ配列から転写されるｍＲＮＡの二次構造に比べて、増大した自由エネルギーを有する。さらに、本方法は修飾型異種ＤＮＡ配列を用いて、組換え発現系において蛋白質を産生する工程を含む。組換え発現系は無細胞試験管内転写翻訳系、試験管内細胞発現系、直接的ＤＮＡ注入に用いられるＤＮＡ構築物、または個体にＤＮＡを導入する為の組換えベクターのいずれでもあり得る。野生型異種ＤＮＡ配列から転写されるｍＲＮＡの二次構造は、コンピューターとコンピュータープログラムとを用いて予測される。野生型異種ＤＮＡ配列はコード配列のアデニンとチミンの含有量を増加することによって修飾されるが、特にコード配列の５’末端、もしくはコード配列の５’末端で、開始コドンから２００ヌクレオチドまたは１５０ヌクレオチドまたは１００ヌクレオチド以内において修飾される。
【０００８】
本発明はまた注射可能な医薬成分に関するものである。本医薬成分は核酸分子を含み、蛋白質をコードする修飾型コード配列が制御可能な形で制御エレメントに結合されている。この修飾型コード配列は野生型核酸配列に比較して、高いアデニンとチミンもしくはアデニンとウラシルの含有量を有している。本医薬成分はさらに医薬的に許容される担体を含む。コードされる蛋白質は免疫原でも非免疫原性治療蛋白質でもよい。コード配列の修飾は最初の１００ないし２００塩基以内、またはコード配列の全体にわたって分散された配列の範囲内、または最後の１００ないし２００塩基以内にされ得る。
【０００９】
本発明はまた、組換えウイルスベクターに関するものである。本ウイルスベクターは核酸分子を含み、蛋白質をコードする修飾型コード配列が制御可能な形で制御エレメントに結合されている。この修飾型コード配列は野生型核酸配列に比較して、高いアデニンとチミンもしくはアデニンとウラシルの含有量を有している。
【００１０】
【発明の実施の形態】
本発明は、ヌクレオチドレベルでの変化に起因する、ＲＮＡの二次構造の自由エネルギーの増加と、二次構造の不安定化によって、蛋白質発現が増強されるという、発見に基づいている。ＲＮＡ転写産物の自由エネルギー（Ｘ　ｋｃａｌ）の増加によって、それがコードする蛋白質発現の増加をもたらすということが知られている。
【００１１】
好ましい実施例において、ＲＮＡ転写産物の２００塩基セグメント範囲内の自由エネルギー（Ｘ　ｋｃａｌ）の増加は、それがコードしている蛋白質発現の増加に帰結するであろう。そのセグメントは、５’末端にあることが好ましく、通常開始コドンを含む。いくつかの実施例においては、そのセグメントの長さが２００塩基、１５０塩基、もしくは１００塩基が好ましい。ＲＮＡ分子の二次構造は、その三次元構造において生じる塩基対の集合体である。任意のＲＮＡ分子の二次構造は予測可能であり、そのように予測された二次構造に対して、総体自由エネルギー価が与えられる。予測されるＲＮＡの二次構造に対して予測される最小限の総体自由エネルギーまで総体自由エネルギーの増加をもたらすような、あるいはまた、そのＲＮＡの部分領域に対して予測される二次構造に対して予測される最小自由エネルギーを増加するような、ＲＮＡ転写産物の一次配列の変更が、そのＲＮＡ転写産物によってコードされる蛋白質の発現の増加を促進することが知られている。
【００１２】
本戦略は蛋白質発現を至適化する為のものであり、発現が望まれるどのような場合にも応

用される。すなわち、生体内におけるＤＮＡワクチン、生ワクチン、遺伝子治療、あるいは遺伝子導入が含まれるが、これらに限定するものではない。また、試験管内においては原核生物や真核生物（哺乳類、昆虫、酵母）の培養細胞の系を用いての組換え製造法を含むが、これらに限定するものではない。また、生体外においては、細胞が発現構築物を受容し、レシピエントの生体に移植される系を含むがこれに限定するものではない。また、目的の遺伝子の発現、もしくは遺伝子発現を増強することが望まれる、他のいかなる発現系も含まれる。

【００１３】

この発明のひとつの側面は、蛋白質の効率的な発現を促進、もしくは蛋白質の発現レベルの増加を導く蛋白質をコードするＲＮＡを産生することである。ＲＮＡの予測二次構造に対する最小総体自由エネルギーの増加を導く、あるいは、ＲＮＡのひとつかそれ以上の領域の予測二次構造に対する最小自由エネルギーを増加するように、ＲＮＡをコードするＤＮＡの配列を変更することによって、コードされた蛋白質の能率の良い、もしくは、より増強した発現を促進する。

【００１４】

ＲＮＡの二次構造の自由エネルギーの増加は、エネルギー最小化に基づいて入力配列に対して最も安定な構造を計算し予測する、ＭＦＯＬＤのようなプログラムを用いて、さまざまに変更されたバージョンの配列を分析することによってモニターすることができる。ＭＦＯＬＤは、ズーカー（Ｚｕｋｅｒ）、ジャガー（Ｊａｅｇｅｒ）、その同僚らによってデザインされたコンピューターソフトウェアであり（ズーカー、１９７８、ＲＮＡ分子の全ての準至適折り畳みについての知見、Ｓｃｉｅｎｃｅ，２４４：４８－５２．ジャガーら、１９８９、ＲＡＮの二次構造の改善された予測、Ｐｒｏｃ．Ｎａｔｌ．Ａｃａｄ．Ｓｃｉ．ＵＳＡ、８６：７７０６－７７１０これらはそれぞれ、ここに参考文献として含まれる）、Ｔｕｒｎｅｒとその同僚らによって開発されたエネルギー法則を用いた自由エネルギー最小化によるＲＮＡ二次構造の予測に用いられる（参考文献としてここに含まれる、Ｆｒｅｉｅｒら、１９８６、Ｐｒｏｃ．Ｎａｔｌ．Ａｃａｄ．Ｓｃｉ．ＵＳＡ、８３：９３７３－９３７７を参照のこと）。ＭｕｌＦｏｌｄはＭＦＯＬＤのマッキントッシュ版である。ＬｏｏｐＤｌｏｏｐは二次構造描写プログラムである。最も安定な構造は最小限の総体自由エネルギーを持つものである。構造に対する自由エネルギー値が負になればなるほど、より安定である。より高い総体自由エネルギー価を持つ二次構造になることが予測されるＲＮＡ配列の変更は、より安定でないＲＮＡの二次構造に帰結し、ＲＮＡの効率的な翻訳と蛋白質発現の増加を促進する不安定な変化である。

【００１５】

この発明の実践は、もし他の方法で指示されなければ、ウイルス学、免疫学、微生物学、分子生物学、組換えＤＮＡ技術で通常の知識に属する在来法を使用する。このような技術は文献の中で十分に説明されている。例えば、サムブルック（Ｓａｍｂｒｏｏｋ）ら、ｅｄｓ．、Ｍｏｌｅｃｕｌａｒ Ｃｌｏｎｉｎｇ：Ａ Ｌａｂｏｒａｔｏｒｙ Ｍａｎｕａｌ（３rd ｅｄ．）Ｃｏｌｄ Ｓｐｒｉｎｇ Ｈａｒｂｏｒ Ｌａｂｏｒａｔｏｒｙ Ｐｒｅｓｓ、Ｃｏｌｄ Ｓｐｒｉｎｇ Ｈａｒｂｏｒ、ＮＹ（２００１）；アースベル（Ａｕｓｕｂｅｌ）ら、ｅｄｓ．、Ｃｕｒｒｅｎｔ Ｐｒｏｔｏｃｏｌｓ ｉｎ Ｍｏｌｅｃｕｌａｒ Ｂｉｏｌｏｇｙ、Ｊｏｈｎ Ｗｉｌｅｙ ＆ Ｓｏｎｓ、Ｎｅｗ Ｙｏｒｋ、ＮＹ（２００１）；Ｇｌｏｖｅｒ ＆ Ｈａｍｅｓ、ｅｄｓ、．ＤＮＡ Ｃｌｏｎｉｎｇ ３：Ａ Ｐｒａｃｔｉｃａｌ Ａｐｐｒｏａｃｈ、Ｖｏｌｓ．Ｉ、ＩＩ ＆ ＩＩＩ、ＩＲＬ Ｐｒｅｓｓ、Ｏｘｆｏｌｄ（１９９５）；Ｃｏｌｏｗｉｃｋ ＆ Ｋａｐｌａｎ、ｅｄｓ．、Ｍｅｔｈｏｄｓ ｉｎ Ｅｎｚｙｍｏｌｏｇｙ、Ａｃａｄｅｍｉｃ Ｐｒｅｓｓ；ウェイアー（Ｗｅｉｒ）ら、Ｈａｎｄｂｏｏｋ ｏｆ Ｅｘｐｅｒｉｍｅｎｔａｌ Ｉｍｍｕｎｏｌｏｇｙ、５th ｅｄ．、Ｂｌａｃｋｗｅｌｌ Ｓｃｉｅｎｔｉｆｉｃ ｐｕｂｌｉｃａｔｉｏｎｓ、Ｌｔｄ．、Ｅｄｉｎｂｕｒｇｈ、（１９９７）；Ｆｉｅｌｄｓ、Ｋｎｉｐｅ、＆Ｈｏｗｌｅｙ、ｅｄｓ．、Ｆｉｅｌｄｓ Ｖｉｒｏｌｏｇｙ（３rd ｅｄ．）Ｖｏｌｓ．Ｉ ＆ ＩＩ、Ｌｉｐｐｉｎｃｏｔｔ Ｗｉｌｌｉ

ａｍｓ　＆　Ｗｉｌｋｉｎｓ　Ｐｕｂｓ．（１９９６）；フリント（Ｆｌｉｎｔ）ら、ｅｄｓ．、Ｐｒｉｎｃｉｐｌｅｓ　ｏｆ　ｖｉｒｏｌｏｇｙ：Ｍｏｌｅｃｕｌａｒ　ｂｉｏｌｏｇｙ、ｐａｔｈｏｇｅｎｅｓｉｓ、ａｎｄＣｏｎｔｒｏｌ、ＡＳＭ　Ｐｒｅｓｓ、（１９９９）；コリガン（Ｃｏｌｉｇａｎ）ら、ｅｄｓ．、Ｃｕｒｒｅｎｔ　Ｐｒｏｔｏｃｏｌｓ　ｉｎ　Ｉｍｍｕｎｏｌｏｇｙ、Ｊｏｈｎ　Ｗｉｌｅｙ　＆　Ｓｏｎｓ、Ｎｅｗ　Ｙｏｒｋ、ＮＹ（２００１）を参照すること。これらの文献はこの明細書に含まれる。

【００１６】

本明細書の全体にわたり、様々な定義がされている。ほとんどの表現は、通常の知識を有する人には通じる言葉である。本明細書内で下記もしくは他のどこかで特に定義される表現は、この発明の文脈の中で通常の知識を有する者によって一般的に理解される。

【００１７】

ここで使われるように、”組換え発現系”という用語は、対象の遺伝子産物（単数または複数）の発現を目的とする、核酸に基づく任意の方法や実験系のことであり、対象の遺伝子産物（単数または複数）はそれらの発現を目的とする成分から人工的に作られてきた（ヒトが作った）ものである。その成分とは、自然に生じた遺伝資源、合成的もしくは人工的、もしくは天然と人工の遺伝的要素の組み合わせである。一般的に遺伝子産物とは蛋白質、ポリペプチド、ペプチドである。組換え発現系の例としては、無細胞試験管内転写翻訳系、試験管内細胞発現系、ＤＮＡ直接注入に用いられるＤＮＡ構築物、そして、個体にＤＮＡを導入する組換えベクターが含まれるが、これらに限定するものではない。

【００１８】

ここで使われるように、”異種ＤＮＡ配列”とは自然に存在しない状況におかれた、デオキシリボ核酸に基づく配列のことであり、これらは例えば、組換え構築物、またはプラスミド、またはウイルスの中にあるか、染色体上の自然に存在しない位置に挿入されているか、もしくは、自然に存在しないかまたは異種の細胞に導入されているものである。”異種ＤＮＡ配列”とは任意のＤＮＡ配列のことであって、それは異種のＤＮＡ配列であるか、またはそれが（制御可能であろうとなかろうと）関連もしくは結合する他のＤＮＡ配列と自然には関連しないＤＮＡ配列であるか、またはそれが導入される細胞や生体と自然には関連しないＤＮＡｎｏ配列のことである。異種ＤＮＡ配列の例としては、宿主細胞または生体において、外来性もしくは異種蛋白質性遺伝子産物の発現に用いられた配列が挙げられる。異種ＤＮＡ配列は、宿主細胞の生体内または試験管内、または無細胞試験管内発現系において、蛋白質、ポリペプチド、またはペプチド等の遺伝子産物を発現するように設計された、遺伝物質を有する、ベクターもしくは発現構築物の一部で有り得る。

【００１９】

ここで使われるように、”野生型異種ＤＮＡ配列”という用語は、非自然環境に位置づけられているが、自然な環境におけるその配列から修飾または変更されていないヌクレオチド配列を持つ異種ＤＮＡ配列のことである。例えば、ウイルス遺伝子は、組換え発現構築物に挿入されて、その構築物内での他の配列との関係において異種ＤＮＡ配列であるが、ウイルス遺伝子の核酸配列にはどんな変化も導入されていない。この例において、野生型異種ＤＮＡ配列として、ウイルス遺伝子はウイルスゲノムの中に自然の状態で存在する通りの野生型核酸配列をもつが、組換え発現構築物（新しい環境）との関係においては異種である。野生型異種ＤＮＡ配列はまた、任意のＤＮＡ配列であって、それは核酸配列の変化を含む修飾型（非自然）ＤＮＡ配列を調整するための、参考配列もしくは出発材料（開始バージョン）となり得る。野生型異種ＤＮＡ配列はまた、自然に存在する配列から変更されていないが、自然界には一緒に存在することがない、野生型ＤＮＡ配列の多数の組み合わせにより成り立つこともあり得る。このように、多数の野生型ＤＮＡ配列から成り立つ野生型異種ＤＮＡ配列の例としては、二つの異なる遺伝子由来の野生型遺伝子配列で構成される融合遺伝子がある。

【００２０】

ここで使われるように、”修飾型異種ＤＮＡ配列”とは、非自然環境に位置付けられるだけでなく、自然環境での配列を修飾もしくは変更されたヌクレオチド配列を持つ異種ＤＮ

Ａ配列のことである。例えば、修飾型異種ＤＮＡ配列であるウイルス遺伝子は、組換え発現構築物に挿入されることによって、構築物の中の他の配列との関係において異種であるが、さらに、ウイルスゲノム中の自然な環境に見つけられる野生型ヌクレオチド配列ではなく、修飾または変更されたヌクレオチド配列を持つ。

【００２１】

ここで使われるように、ＲＮＡ二次構造に関する”自由エネルギーの増加”という用語は、ＲＮＡ二次構造に対する自由エネルギー価の増加のことを示す。より多く負である自由エネルギー価は、より少なく負であるところの自由エネルギー価よりも低い。

【００２２】

ここで使われるように、”修飾型コード配列”とは、核酸配列（ＤＮＡ－またはＲＮＡ－に基づく）のことであり、それは遺伝子産物、蛋白質、ポリペプチド、もしくはペプチドをコードし、しかも、野生型または自然に生じるその遺伝子産物、蛋白質、ポリペプチド、もしくはペプチドに対するコード配列から修飾、変更されたものである。そのコード配列はひとつ以上の遺伝資源由来の配列で構成され得る。例えば、そのコード配列は、ある蛋白質の遺伝子由来のリーダー配列と、それ以外の部分は別の蛋白質の遺伝子由来の配列とから成り立つ融合蛋白質をコードする融合遺伝子で有り得る。この融合遺伝子は、一つの融合コード配列として一緒に運ばれ、自然には存在しないものである。この例ように、ひとつ以上の遺伝資源由来の複数の配列から成り立つコード配列の場合、”修飾型コード配列”とは、いかなる修飾も、それぞれ別個のコード配列について、野生型もしくは自然に生じるコード配列に関連して行われたものであることを示す。

【００２３】

ここで使われるように、”野生型コード配列”という用語は、核酸配列（ＤＮＡ－またはＲＮＡ－に基づく）のことであり、それは遺伝子産物、蛋白質、ポリペプチド、もしくはペプチドをコードし、しかも、野生型または自然に生じるその遺伝子産物、蛋白質、ポリペプチド、もしくはペプチドに対するコード配列から修飾、変更されていないものである。コード配列が融合蛋白質をコードする場合、成分部分は、それら構成部分に対する野生型または自然に生じるコード配列から修飾もしくは変更されていない。

【００２４】

ここで使われるように、”より高いアデニンとチミンもしくはアデニンとウラシル含有量”という用語は、コード配列に対する修飾のことであり、その修飾はコード配列を、もしそれがＤＮＡに基づいたものであれば、対応する野生型コード配列より高い濃度のアデニンとチミジン残基を持ち、もしＲＮＡに基づいたものであれば、対応する野生型コード配列より高い濃度のアデニンとウリジン残基を持つ、というような、”修飾コード配列”にすることを指す。

【００２５】

ここで使われるように、野生型コード配列と比較して、修飾型コード配列に関しての”最初の２００塩基”という用語は、それぞれのコード配列の５’末端からの最初の隣接２００ヌクレオチド基を示す。

【００２６】

ここで使われるように、野生型コード配列と比較して、修飾型コード配列に関しての”最初の１５０塩基”という用語は、それぞれのコード配列の５’末端からの最初の隣接１５０ヌクレオチド基を示す。

【００２７】

ここで使われるように、野生型コード配列と比較して、修飾型コード配列に関しての”最初の１００塩基”という用語は、それぞれのコード配列の５’末端からの最初の隣接１００ヌクレオチド基を示す。

【００２８】

ここで使われるように、野生型コード配列と比較して、修飾型コード配列に関しての”最後の２００塩基”という用語は、それぞれのコード配列の３’末端からの最後の隣接２００ヌクレオチド基を示す。

【００２９】
ここで使われるように、野生型コード配列と比較して、修飾型コード配列に関しての”最後の１５０塩基”という用語は、それぞれのコード配列の３’末端からの最後の隣接１５０ヌクレオチド基を示す。
【００３０】
ここで使われるように、野生型コード配列と比較して、修飾型コード配列に関しての”最後の１００塩基”という用語は、それぞれのコード配列の３’末端からの最後の隣接１００ヌクレオチド基を示す。
【００３１】
ここで使われるように、コード配列に関しての”２００塩基の長さまでの領域”という用語は、コード配列の隣接２００ヌクレオチド基までの領域を示す。
【００３２】
ここで使われるように、コード配列に関しての”１５０塩基の長さまでの領域”という用語は、コード配列の隣接１５０ヌクレオチド基までの領域を示す。
【００３３】
ここで使われるように、コード配列に関しての”１００塩基の長さまでの領域”という用語は、コード配列の隣接１００ヌクレオチド基までの領域を示す。
【００３４】
ここで使われるように、”分散修飾”という用語は、修飾を受ける個々の領域が野生型コード配列に比べて、より高いアデニンとチミンもしくはアデニンとウラシル含有量を有するように修飾される、隣接したヌクレオチド基群の少なくとも二つの領域の任意の組み合わせのことを示し、しかも、修飾型コード配列の領域が野生型コード配列と交互に配置されるように、配列の全般に分散されることを示す。これに限るものではないが、例として、修飾型コード配列は修飾された配列領域を交互に含み得る。そして、コード配列の最初の隣接した２００塩基は野生型コード配列と比較してより高いアデニンとチミンもしくはアデニンとウラシルの含有量を持っており、コード配列の次の２００塩基は野生型コード配列に比較して非修飾配列であり、後の隣接した２００塩基領域は野生型コード配列と比較してより高いＡＴまたはＡＵ含有量を持つように修飾される、という具合である。修飾領域の長さは任意であるが、２００、１５０、１００塩基の長さが好ましい。非修飾領域の長さは、修飾領域の位置に依存するので、さまざまな長さになる。好ましい実施例として、分散修飾は、コード配列の全長にわたり、修飾型コード配列と野生型コード配列の交互の領域を有しており、その交互の領域のサイズはそれぞれ、２００塩基または１５０塩基または１００塩基の長さが好ましい。
【００３５】
ここで使われている、”注射可能な医薬品成分”という用語は、患者における使用に対して、医薬品として容認可能な成分であるということ示す。したがって、無菌であり、発熱物質を含まず、本質的に微粒子や微粒物質を含まないものである。Ｒｅｍｉｎｇｔｏｎ’ｓ Ｐｈａｒｍａｃｅｕｔｉｃａｌ Ｓｃｉｅｎｃｅｓ、１８th Ｅｄ．、Ｇｅｎｎａｒｏ、ｅｄ．、Ｍａｃｋ Ｐｕｂｌｉｓｈｉｎｇ Ｃｏ．、Ｅａｓｔｏｎ、ＰＡ、１９９０とＵ．Ｓ．Ｐ．、Ｕ．Ｓ．薬局方の基礎、をここに参照として記載する。
【００３６】
ここで使われている”医薬品として容認可能な担体”は、成分を受取った個人に有害な影響を誘発しない、さまざまな担体のことを含んでいる。例えば、”医薬品的に容認可能な担体”はレシピエントに対して有害な抗体の産生を誘発すべきでない。適切な”医薬品的に容認可能な担体”は、通常の知識を有している者には知られており、Ｒｅｍｉｎｇｔｏｎ’ｓ Ｐｈａｒｍａｃｅｕｔｉｃａｌ Ｓｃｉｅｎｃｅｓ、ｓｕｐｒａ。に記載されている。
【００３７】
ここで使われている、”標的蛋白質”という用語は、免疫反応に対する標的蛋白質として働き、この発明における遺伝子構築物によってコードされているペプチドと蛋白質を示す

。”標的蛋白質”や”免疫原”という用語は交換可能で、それに対して免疫反応を誘発させることができる蛋白質を言及する。標的蛋白質とは免疫原性の蛋白質であり、それは病原体または癌細胞や自己免疫疾患に関係する細胞のような望ましくないタイプの細胞由来の蛋白質と少なくともひとつの抗原決定基を共有しており、それに対する免疫反応が望まれるもののことである。標的蛋白質に対して惹起される免疫反応は、標的蛋白質が関係する、特定の感染症や疾患から個体を保護し、または治療に有用であろう。

【００３８】

ここで使われている、”望まれる蛋白質”は、本発明における遺伝子構築物によってコードされるペプチドと蛋白質のことであり、その遺伝子構築物は免疫反応に対する標的蛋白質として、もしくは遺伝子療法における治療的、補償的蛋白質としても働く。

【００３９】

ここで使われている、免疫原に関する”その免疫原性断片”という句は、それに対して免疫反応が惹起される、免疫原の全長より短い断片のことである。

【００４０】

ここで使われている、”癌抗原”という用語は、癌、腫瘍、もしくは癌細胞に関係し、またはそのマーカーとして働く、任意の蛋白質、ポリペプチドもしくはペプチド、あるいは類似の物質を示す。

【００４１】

ここで使われている、”自己免疫疾患関連蛋白質”とは任意の蛋白質、ポリペプチド、またはペプチド、ならびに類似の物質であって、自己免疫疾患に関与もしくはその原因となる細胞に関連するか、またはその細胞のマーカーとなるものである。

【００４２】

ここで使われている、”非免疫原性治療蛋白質”という用語は、さまざまな病気や異常の治療法において有用な蛋白質、ポリペプチド、ペプチドを示し、そのような治療や処置が必要であるレシピエントの生体、患者もしくは個人の身体への導入で、免疫反応は望まれず、またはもしくは予測されないものを指す。”非免疫原性治療蛋白質”の例として、その蛋白質をコードしている内因性遺伝子の遺伝的欠陥を持つ個人において、欠損もしくは低濃度である蛋白質が挙げられる。”非免疫原性治療蛋白質”の例として、限定するものではないが、サイトカイン、成長因子、血液産物、酵素が含まれる。

【００４３】

ここで使われている、”組換えウイルスベクター”という用語は、ウイルスゲノムに基づいた構築物を示し、それは、対象の蛋白質、ポリペプチド、ペプチドをコードする核酸を運ぶための担体として使われる。組換えウイルスベクターは通常の知識を有するところにおいてはよく知られており、広く報告されている。組換えウイルスベクターは、限定するものではないが、レトロウイルスベクター、アデノウイルスベクター、アデノ関連ウイルスベクターを含み、それらは普通の方法と出発材料を用いて準備される。

【００４４】

ここで使われている、”遺伝的構築物”という用語は、標的蛋白質または免疫修飾蛋白質をコードするヌクレオチド配列を有するＤＮＡまたはＲＮＡ分子のことである。そのコード配列は、核酸分子を投与された個々の細胞において直接発現できるように、プロモーターとポリアデニル化シグナルを含む調節エレメントと連結可能な開始、終結シグナルを含む。

【００４５】

ここで使われている、”発現可能な形体”という用語は、遺伝子構築物のことである。これは必須調節エレメントを持ち、標的蛋白質または免疫修飾蛋白質をコードするコード配列に制御可能な形で連結されており、個々の細胞内に存在する時、そのコード配列が発現される。

【００４６】

ここで使われている、”抗原決定基の共有”という用語は、他の蛋白質の抗原決定基と同一もしくはほぼ同じである、少なくともひとつの抗原決定基を有する蛋白質を示す。

【００４７】
ここで使われている、”実質的に同様の抗原決定基”という用語は、蛋白質の抗原決定基と同一ではないが、それにも関わらず、その蛋白質と交差反応する細胞性もしくは体液性免疫反応を惹起する構造を持つ抗原決定基のことを意味する。
【００４８】
ここで使われている、”細胞内病原体”という用語は、複製または生活環の少なくとも一区間で宿主細胞の中に存在し、その中で病原体蛋白質を生産もしくは生産を引き起こす、ウイルスもしくは病原生物体のことを意味する。
【００４９】
ここで使われている、”過剰増殖病”という用語は、細胞の過剰増殖を特徴とする病気や異常のことを意味している。
【００５０】
ここで使われている、”過剰増殖関連蛋白質”という用語は、過剰増殖病、およびもしくは過剰増殖細胞に関連した蛋白質のことを意味している。
【００５１】
いくつかの好ましい実施例において、ＲＮＡの変化は蛋白質の配列を変化しないことが好まれる。いくつかの好ましい実施例において、変化が導入された２００塩基は、ＲＮＡ転写産物の５’末端にあることが好まれる。いくつかの好ましい実施例において、ＲＮＡ転写産物のひとつ以上のセグメントにおいて自由エネルギーが増加することが好まれる。ＲＮＡ二次構造の自由エネルギー増加のために、任意にリーダー配列を加える。
【００５２】
オープンリーディングフレーム（ｏｒｆ）配列の５’末端にある安定性ＲＮＡ二次構造は、リボゾーム機能を阻害することによって転写を効果的にブロックする。多くのＲＮＡ分子は高度に安定な二次構造の整合性を持っており、これらの相互作用は、遺伝子発現を阻害できる。アデニンとチミンが豊富な配列で至適化したように修飾型、リーダー配列をコードする配列の添加は、予測ＲＮＡ構造のより高い自由エネルギーに帰結し、細胞性リボゾームによる効果的な開始を認めた。安定性ＲＮＡ二次構造は自由エネルギーの増加によって除かれる。
【００５３】
従って、この発明によれば、ＲＮＡ二次構造の整合性の減少と、ＲＮＡ転写産物における阻害二次構造（ステムループ）の解離による蛋白質の発現／翻訳の増加という結果によって、コード配列内のアデニンとウラシルの含有量の増加は配列を至適化する。二次構造の整合性の分解は、ＲＮＡやコード配列、特にそのＲＮＡの最初の１００から２００ヌクレオチド、の５’部分において特に大事である。いくつかの実施例において、転写開始点から最初の１００から２００ヌクレオチドにおいてアデニンとウラシルのもしくはアデニンとチミンの含有量が増加し、いくつかの具体化の中においては、コード配列または翻訳開始点の最初の１００から２００ヌクレオチドにおいてアデニンとウラシルもしくはアデニンとチミンの含有量が増加する。いくつかの実施例において、ＲＮＡ二次構造完全体の破損は、全遺伝子変化もしくは１００から２００ヌクレオチド基伸張における変化パターンによって成立する。３’末端の修飾も大事である。
【００５４】
リーダーコード配列の添加や、酵母至適化（滅多にヒトのコドンは用いられない）されたその配列のコドン変化の戦略は、どんな蛋白質をコードしたどんな遺伝子にも適応できる。限定するものではないが、例えば、ＨＩＶ－１　ｐｏｌ遺伝子を含むウイルス蛋白質をコードする遺伝子が挙げられる。アデニンとウラシルの含有量が豊富であることがより好ましく；ヒト優性コドン／グアニンとシトシンの含有量が高いものはあまり好ましくない。蛋白質発現や構造に関する前もった知識なしでｍＲＮＡでの二次構造の領域の安定性を低下可能である、ということが発見されていた。
【００５５】
結果として生じる変化した転写産物から予測された二次構造の最小自由エネルギーは、元

の転写産物よりも増強した蛋白質発現が可能である変化した転写産物を与える。標準的な技術と容易に入手可能な出発材料を用いて、修飾核酸分子を準備する。その核酸分子は、宿主細胞に取り込まれる発現ベクターに組み込まれる。蛋白質産生に対してよく知られた組換え発現系に用いる宿主細胞は、よく知られており容易に入手可能である。宿主細胞の例としては、Ｅ．ｃｏｌｉ（大腸菌）のような細菌細胞、Ｓ．ｃｅｒｅｖｉｓｉａｅのような酵母細胞、Ｓ．ｆｒｕｇｉｐｅｒｄａのような昆虫細胞、チャイニーズハムスター卵巣（ＣＨＯ）細胞などの非ヒト哺乳類組織培養細胞、ＨｅＬａ細胞のようなヒト組織培養細胞が含まれる。

【００５６】

いくつかの実施例において、例えば、通常の知識を有してる人は、よく知られた技術を用い、よく知られた発現系として用いるための市販の発現ベクターに、ＤＮＡ分子を挿入できる。例えば、市販のプラスミドであるｐＳＥ４２０（Ｉｎｖｉｔｒｏｇｅｎ、Ｓａｎ Ｄｉｅｇｏ、ＣＡ）は、Ｅ．ｃｏｌｉにおける免疫修飾蛋白質の産生に用いられる。また、例えば、市販のプラスミドｐＹＥＳ２（Ｉｎｖｉｔｒｏｇｅｎ、Ｓａｎ Ｄｉｅｇｏ、ＣＡ）は、酵母のＳ．ｃｅｒｅｖｉｓｉａｅにおける産生に用いられる。例えば、市販のＭＡＸＢＡＣ^{TM}コンプリートバキュロウイルス発現系（Ｉｎｖｉｔｒｏｇｅｎ、Ｓａｎ Ｄｉｅｇｏ、ＣＡ）は、昆虫細胞における産生に用いられる。市販のプラスミドｐｃＤＮＡＩもしくはｐｃＤＮＡ３（Ｉｎｖｉｔｒｏｇｅｎ、Ｓａｎ Ｄｉｅｇｏ、ＣＡ）は、ＣＨＯ細胞のような哺乳類細胞における産生に用いられる。通常の知識を有する人は、これらの市販の発現ベクターや発現系、もしくは、普通の技術や容易に利用可能な出発材料による免疫修飾蛋白質を産生するための他の技術を用いることができる。（サムブルックら、ｅｄｓ．、２００１、ｓｕｐｒａ参照）従って、要求蛋白質は、原核生物、真核生物の両方の系において準備でき、その結果、さまざまな形の処理された蛋白質が生じる。

【００５７】

通常の知識を有する人は、市販の他の発現ベクターや系を用いることもあるし、もしくはよく知られた方法や容易に利用可能な出発材料を用いてベクターを作製する。プロモーターやポリアデニル化シグナルのような必須調節配列やエンハンサーを含む発現系は、容易に利用可能で、宿主の多様性に対する技術として知られている。（サムブルックら、ｅｄｓ．、２００１、ｓｕｐｒａ参照）。

【００５８】

修飾ＤＮＡを含む発現ベクターは、互換性のある宿主を形質転換するために用いられ、その後、外来ＤＮＡの発現が起こる条件下で培養され、維持される。この発明において産生される蛋白質は、通常の知識を有する者に適切でよく知られているように、培養によって、細胞の溶解によって、または培養液中に回収される。通常の知識を有した人は、よく知られている技術を使い、その発現系を用いて産生された蛋白質を分離することができる。抗体を用いて自然源から蛋白質を精製する方法は、組換えＤＮＡ方法論によって生産された精製蛋白質に同様に当てはめられる。

【００５９】

この発明の医薬品成分は、哺乳類の身体の中において活性を持った薬剤がその作用場所へ到達することを可能にする、任意の手段によって投与される。この発明の医薬品成分は、局所または全身の治療が必要とされているかということと、治療する場所に依存した多数の方法によって投与される。投与は、局所的（眼、膣、直腸、鼻腔内、経皮を含む）、経口、もしくは非経口投与である。非経口投与は、静脈点滴、皮下・腹腔内・筋肉内注射、例えば吸引やガス注射による肺投与、くも膜下腔内または脳室内投与が含まれる。

【００６０】

この発明は核酸分子を有する注射可能な医薬品成分と関係している。

【００６１】

調節エレメントと連結可能な修飾ヌクレオチド配列を有する注射可能な医薬品成分は、ＤＮＡ注射（ＤＮＡワクチン接種としても言及される）、組換えアデノウイルス、ウイルス関連組換えアデノウイルス、そして組換えワクシニアウイルスのような組換えベクターを

含めて、いくつかのよく知られた技術のどれかを用いて運ばれる。

【００６２】

ＤＮＡワクチンは米国特許番号．５，５９３，９７２、５，７３９，１１８、５，８１７，６３７、５，８３０，８７６、５，９６２，４２８、５，９８１，５０５、５，５８０，８５９、５，７０３，０５５、５，６７６，５９４、とここで引用されている前明細書に記載されている。それらの明細書に記載されている添付実験方法に付け加えて、導入ＤＮＡの選択方法は、ここの参照に含まれている、米国特許番号４，９４５，０５０と５，０３６，００６、に記載されている。投与方法は、限定するものではないが、吸引、坐薬によって、もしくは粘膜組織、例えば膣、直腸、尿道、頬側、舌下組織における洗浄によって、筋肉内、鼻腔内、腹腔内、皮内、皮下、静脈内、動脈内、眼内、経口、また、局所的、経皮的投与を含む。より好ましい投与方法は、粘膜組織、筋肉内、腹腔内、皮内、皮下注射を含む。遺伝子構築物は、限定するものではないが、従来の注射針のない注射装置、もしくは”マイクロプロジェクティルボンバードメントジーンガン”を含む手段によって投与される。

【００６３】

細胞を取り扱う時、遺伝子構築物は、機能する染色体外分子として細胞内に残り、あるいは細胞の染色体ＤＮＡの中に結合する。ＤＮＡはプラスミドもしくはプラスミドらの形で分離遺伝子物質として維持し細胞に導入される。代わりに、染色体内に結合できるリニアＤＮＡが細胞に導入される。細胞内にＤＮＡを導入する時、ＤＮＡが染色体に結合することを促進する試薬が加えられる。結合を促進するのに有効なＤＮＡ配列もＤＮＡ分子に含まれる。代わりに、ＲＮＡが細胞に投与される。セントロメア、テロメア、複製基点を含むリニアミニ染色体としての遺伝子構築物を供給することも予測される。遺伝子構築物は、減弱した生微生物もしくは細胞内で生存している組換え微生物ベクター内に、遺伝子物質の一部を残している。遺伝子構築物は組換えウイルスワクチンのゲノムの一部分で有り得、その中の遺伝物質は細胞の染色体に組み込まれるか、もしくは染色体外に残ることになる。

【００６４】

遺伝子構築物は核酸分子の遺伝子発現に必要な調節エレメントを含む。そのエレメントには、プロモーター、開始コドン、ストップコドン、ポリアデニル化シグナルを含まれる。付け加えると、エンハンサーはしばしば、標的蛋白質をコードする配列の遺伝子発現に必要とされる。これらの因子は要求蛋白質をコードする配列と連結可能であり、調節エレメントは投与された個体において作用可能である、ということが必要である。開始コドンとストップコドンは一般的に要求蛋白質をコードするヌクレオチド配列の一部と考えられている。しかしながら、これらの因子は、その遺伝子構築物を投与された個々において機能的であるということが必要である。その開始と終結コドンはコード配列の枠内でなくてはならない。プロモーターとポリアデニル化シグナルは個々の細胞内において機能的でなくてはならない。この発明の実施に対して有効なプロモーターの例として、特にヒトに対する遺伝子ワクチンの産生においては、ヒトアクチン、ヒトミオシン、ヒトヘモグロビン、ヒト筋肉クレアチン、ヒトメタロチオネインのようなヒト遺伝子由来のプロモーター、Ｓｉｍｉａｎ Ｖｉｒｕｓ ４０（ＳＶ４０）プロモーター、Ｍｏｕｓｅ Ｍａｍｍａｒｙ Ｔｕｍｏｒ Ｖｉｒｕｓ（ＭＭＴＶ）プロモーター、Ｈｕｍａｎ Ｉｍｍｕｎｏｄｅｆｉｃｉｅｎｃｙ Ｖｉｒｕｓ（ＨＩＶ）のＨＩＶ Ｌｏｎｇ Ｔｅｒｍｉｎａｌ Ｒｅｐｅａｔ（ＬＴＲ）プロモーター、Ｍｏｌｏｎｅｙ ｖｉｒｕｓプロモーター、ＡＬＶのプロモーター、Ｃｙｔｏｍｅｇａｌｏｖｉｒｕｓ（ＣＭＶ）のＣＭＶ前初期プロモーター、Ｅｐｓｔｅｉｎ Ｂａｒｒ Ｖｉｒｕｓ（ＥＢＶ）のプロモーター、Ｒｏｕｓ Ｓａｒｃｏｍａ Ｖｉｒｕｓ（ＲＳＶ）のプロモーターを含むが、これらに限定するものではない。この発明の実施に有効であるポリアデニル化シグナルの例は、特にヒトに対する遺伝子ワクチンの産生において、限定するものではないが、ＳＶ４０ポリアデニル化シグナルとＬＴＲポリアデニル化シグナルが含まれる。特に、ＳＶ４０ポリアデニル化シグナルとして言及されているｐＣＥＰ４プラスミド（Ｉｎｖｉｔｒｏｇｅｎ、Ｓａｎ Ｄｉｅｇｏ、

ＣＡ）の中のＳＶ４０ポリアデニル化シグナルが使われている。ＤＮＡ発現に必要な調節エレメントに付け加えて、他の因子はＤＮＡ分子の中に含まれる。そのような追加因子はエンハンサーを含む。そのエンハンサーは、限定するものではないが、ＣＭＶ、ＲＳＶ、ＥＢＶなどのウイルス由来のエンハンサー、ならびにヒトアクチン、ヒトミオシン、ヒトヘモグロビン、ヒト筋肉クレアチンのエンハンサーを含むグループから選ばれるものである。構築物を染色体外性に維持し、細胞内に多数の複製物を産生するために、遺伝子構築物には哺乳類複製基点を加えることが可能である。Ｉｎｖｉｔｒｏｇｅｎ（Ｓａｎ Ｄｉｅｇｏ、ＣＡ）製プラスミドｐＣＥＰ４とｐＲＥＰ４は、Ｅｐｉｓｔｅｉｎ Ｂａｒｒ ｖｉｒｕｓの複製基点ならびに核抗原ＥＢＮＡ－１コード領域を含んでおり、染色体に組み込まれることなしに、染色体外で高コピー数の複製が行われる。

【００６５】

この発明のひとつの方法は、筋肉内、鼻腔内、腹腔内、皮下、皮内、もしくは局所、の核酸分子の投与ステップ、もしくは吸引、膣、直腸、尿道、頬側、舌下から成るグループから選択された粘膜組織の洗浄によるステップを有する。

【００６６】

いくつかの実施例において、核酸分子はポリヌクレオチド機能エンハンサー、または遺伝子ワクチン促進因子の投与と協力して働く細胞に運ばれる。ポリヌクレオチド機能エンハンサーは１９９３年１月２６日に提出された米国特許出願番号０８／００８，３４２、１９９３年3月１１日に提出された米国特許出願番号０８／０２９，３３６、１９９３年９月２１日に提出された米国特許出願番号０８／１２５，０１２、１９９４年1月２６日に提出された国際出願番号ＰＣＴ／ＵＳ９４／００８９９に記載されており、それぞれはここの参照として記載してある。遺伝子ワクチン促進（ＧＶＦ）因子は、ここの参照として含まれている、１９９４年４月１日に提出された米国特許出願番号０８／２２１，５７９に記載されている。核酸分子と協力して働く投与される共因子は、核酸分子の投与の前か後に核酸分子との混合として投与され、もしくは別々の形で同時に投与される。さらに、因子ＧＶＦと共に共投与されるであろう他の因子で、形質移入因子、または複製因子、または、炎症性因子として働くものの中には、成長因子、サイトカイン、リンホカインを含み、それにはα－インターフェロン、γ－インターフェロン、血小板由来成長因子（ＰＤＧＦ）、ＴＮＦ、上皮細胞成長因子（ＥＧＦ）、ＩＬ－１、ＩＬ－２、ＩＬ－４、ＩＬ－６、ＩＬ－１０、ＩＬ－１２、線維芽細胞成長因子、免疫刺激複合体（ＩＳＣＯＭＳ）のような界面活性剤、フレウンズ不完全アジュバンド、モノホスホリルリピッドＡ（ＭＰＬ）、ムラミルペプチド、キノン類縁体、スクアレインとスクアレンのような空胞とヒアルロン酸含むＬＰＳ類縁体のようなものがあり、これらはは遺伝子構築物と共に働いて投与される。いくつかの実施例において、免疫調節蛋白質はＧＶＦとして使われる。

【００６７】

この発明によれば、医薬品成分は約１ナノグラムから約２０００マイクログラムのＤＮＡを有する。いくつかの好ましい実施例において、この発明によれば、医薬品成分は約５ナノグラムから約１０００マイクログラムのＤＮＡを有する。いくつかの好ましい実施例においてその医薬品成分は約１０ナノグラムから約８００マイクログラムのＤＮＡを含む。いくつかの好ましい実施例においてその医薬品成分は約０．１から約５００マイクログラムのＤＮＡを含む。いくつかの好ましい実施例においてその医薬成分は約１から約３５０マイクログラムのＤＮＡを含む。いくつかの好ましい実施例においてその医薬品成分は約２５から約２５０マイクログラムのＤＮＡを含む。いくつかの好ましい実施例においてその医薬品成分は約１００から約２００マイクログラムのＤＮＡを含む。

【００６８】

この発明によれば、医薬品成分は使われる投与方法に従って処方される。医薬品成分は注射用薬剤構築物である場合、無菌であり、発熱物質や粒子を含まない。等張処方の使用がより好ましい。一般的に等張性に対する添加物は、塩化ナトリウム、ブドウ糖、マンニトール、ソルビトール、ラクトースを含むことができる。いくつかの場合において、リン酸緩衝食塩水のような等張性水溶液が用意された。安定剤は、ゼラチン、アルブミンを含む

。いくつかの実施例において、血管収縮因子が処方に加えられる。

【００６９】

この発明は、病原体、アレルゲン、もしくは個人の所有する”異常”細胞に特に関連した標的蛋白質に対する広範囲の免疫反応を誘発するために有用である。病原体蛋白質に対する免疫反応はその病原体に対する防衛免疫を供給するように、この発明は病原体因子と有機体に対して個人を免疫化するために有用である。過剰増殖細胞に特に関連する標的蛋白質に対する免疫反応の誘発による癌のような過剰増殖病や障害と戦うのに、この発明は有用である。自己免疫状態に関わる細胞に特に関連する標的細胞に対する免疫反応の誘発による自己免疫疾患や障害に、この発明は有用である。

【００７０】

核酸分子はプラスミドＤＮＡ、組換えベクターの核酸分子として、もしくは弱毒ワクチンまたは細胞ワクチン内に提供される遺伝子物質の一部として供給される。一方いくつかの実施例において、標的蛋白質、およびもしくは免疫調節蛋白質は、それらをコードする核酸分子を添加する蛋白質、もしくはそれらをコードする核酸分子の代わりの蛋白質として運ばれる。

【００７１】

この発明は、単細胞病原組織と多細胞寄生虫のような、ウイルス、原核生物、病原体真核生物組織などの全ての病原体に対して個人を免疫するのに有用である。この発明は、細胞を感染させ、淋病、リステリア、赤痢菌のようなウイルス、原核生物などの被包性ではないそれら病原体に対して個人を免疫するのに特に有用である。付け加えるとこの発明は、細胞内病原体となる生活環段階を含む原生動物病原体に対する個人を免疫するのに有用である。

【００７２】

病原体の感染を防ぐための遺伝子ワクチンを産生するために、免疫防御反応の攻撃に対抗する免疫原性蛋白質をコードする遺伝子物質は、標的蛋白質に対するコード配列として遺伝子構築物に含まれなくてはならない。この発明において特に有用である、細胞内に感染する病原体に対してであろうと、細胞外に感染する病原体に対してであろうと、全ての病原抗原は防御反応を誘発する可能性は低い。なぜならＤＮＡとＲＮＡは両方とも比較的小さく、比較的簡単に産生され得るので、この発明は、多数の病原抗原のワクチン接種を念頭に入れた付加的な利点を供給する。遺伝子ワクチンで用いられた遺伝子構築物は、たくさんの病原抗原をコードする遺伝子物質を含むことができる。例えば、いくつかのウイルス遺伝子は、そこからの複数標的を供給するひとつの構築物に含まれる。

【００７３】

この発明の他の側面は、過剰増殖病に特徴的な過剰増殖細胞に対する広範囲の免疫防御反応を与える方法および過剰増殖病に苦しむ個人の治療法を供給する。過剰増殖病の例として、全ての形態の癌と乾癬を含む。

【００７４】

免疫原性”過剰増殖細胞”関連蛋白質をコードするヌクレオチド配列を含む遺伝子構築物の個々の細胞への導入は、個々のワクチン接種された細胞内のそれらの蛋白質産生に帰結する。過剰増殖病に対して免疫化するために、過剰増殖病に関連する蛋白質をコードするヌクレオチド配列を含む遺伝子構築物が、個々に投与される。

【００７５】

過剰増殖関連蛋白質が効果的な免疫原性標的となるために、過剰増殖細胞において正常細胞に比べて独占的にもしくはより高いレベルで産生される蛋白質であるべきである。標的抗原は、そのような蛋白質から見つけられる少なくともひとつのエピトープを有する蛋白質、フラグメント、ペプチドを含んでいる。いくつかの場合において、過剰増殖関連蛋白質は蛋白質をコードする遺伝子の変異産物である。結果として正常蛋白質には見つけられない異なるエピトープを生じる、わずかに異なったアミノ酸配列を持つこと以外は、その変異遺伝子は正常蛋白質とほぼ同一である蛋白質をコードする。そのような標的蛋白質は癌遺伝子によってコードされている、ｍｙｂ、ｍｙｃ、ｆｙｎ、転位遺伝子であるｂｃｒ

／ａｂｌ、ｒａｓ、ｓｒｃ、Ｐ５３、ｎｅｕ、ｔｒｋ、ＥＧＲＦを含む。標的抗原としての癌遺伝子産物に加えて、抗ガン治療と防御療法に対する標的蛋白質は、Ｂ細胞リンパ腫によって作られた抗体の可変領域と、自己免疫疾患に対する標的抗原として使われるＴ細胞リンパ腫のＴ細胞受容体の可変領域を含んでいる。他の癌関連蛋白質は、モノクローナル抗体１７－１Ａ、葉酸結合蛋白質によって認識される蛋白質を含む、癌細胞に高いレベルで見られる蛋白質のような標的蛋白質として使われる。

【００７６】

癌のいくつかの形態のひとつもしくはそれ以上に対して免疫するために使われている一方、この発明は、特に癌が発症しやすい、もしくは癌患者で再発しやすい個人に、予防的に免疫する時に特に有用である。免疫学と同様に遺伝子学と科学技術における発展によって、個人における癌の発症に対する確率とリスク評価の決定を考慮できる。遺伝子スクリーニングおよびまたは家族健康歴用いることで、特定の個人が癌のいくつかのタイプのどのひとつが発症するかという確率を予測することが可能である。

【００７７】

同様に、すでに発症した癌患者と癌を除去する治療を受けた、もしくは別に緩解にあるそれらの個人は、特に再発と再発現しやすい。治療法の一部として、そのような個人は、診断されたその癌に対して免疫され得る。従って、癌タイプを持っており再発の危険性がある個人は、癌になる将来と戦う彼ら個人の免疫系を準備するために免疫化され得る。

【００７８】

この発明は過剰増殖疾患を患った個人を治療する方法を供給する。この方法において、遺伝子構築物の導入は、標的蛋白質を産生する過剰増殖細胞と戦うように個人の免疫系を仕向けるまたは促進する免疫療法として貢献する。

【００７９】

この発明は、”自己”攻撃抗体を産生する細胞受容体と細胞を含む、自己免疫に関連した標的に対する広い保護免疫反応を与えることによって、自己免疫疾患または障害を患った個人を治療する方法を供給する。

【００８０】

Ｔ細胞は、関節リウマチ（ＲＡ）、多発性硬化症（ＭＳ）、シェーグレン症候群、サルコイドーシス（類肉腫症）、インスリン依存性糖尿病（ＩＤＤＭ）、自己免疫甲状腺炎、反応性関節炎、強直性脊椎炎、強皮症、多発性筋炎、皮膚筋炎、乾癬、血管炎、ウェグナー肉芽腫症、クローン病、潰瘍性大腸炎を含む自己免疫病を媒介した。これらの疾患は、内因性抗原と結合し、自己免疫疾患に関連する炎症性のカスケードを開始するＴ細胞受容体によって特徴付けられる。Ｔ細胞の様々な領域に対するワクチン接種は、それらのＴ細胞を排除するためのＣＴＬｓを含む免疫反応を誘発する。

【００８１】

ＲＡ（関節リュウマチ）において、その病気に関わるＴ細胞受容体（ＴＣＲｓ）のいくつかの特定可変領域は特徴付けられていた。これらのＴＣＲｓはＶβ－３、Ｖβ－１４、Ｖβ－１７、Ｖα－１７を含んでいる。従って、これらの蛋白質の少なくともひとつをコードしているＤＮＡ構築物のワクチン接種は、ＲＡに関わるＴ細胞を標的とする免疫反応を誘発する。ここの参照に取り込まれている、ホウエル（Ｈｏｗｅｌｌ）ら、１９９１、Ｐｒｏｃ．Ｎａｔｌ．Ａｃａｄ．Ｓｃｉ．ＵＳＡ、８８：１０９２１－１０９２５；パリアード（Ｐａｌｉａｒｄ）ら、１９９１、Ｓｃｉｅｎｃｅ、２５３：３２５－３２９；ウィリアムスら、１９９２、Ｊ．Ｃｌｉｎ．Ｉｎｖｅｓｔ．、９０：３２６－３３３；を見ること。

【００８２】

ＭＳ（多発性硬化症）において、その病気に関わるＴＡＲｓのいくつかの特定可変領域は特徴付けられていた。これらのＴＣＲｓはＶβ－７、Ｖα－１０を含む。従って、これらの蛋白質の少なくともひとつをコードしているＤＮＡ構築物のワクチン接種は、ＭＳに関わるＴ細胞を標的とする免疫反応を誘発する。ここの参照に含まれている、ウチェルプフェニグ（Ｗｕｃｈｅｒｐｆｅｎｎｉｇ）ら、１９９０、Ｓｃｉｅｎｃｅ、２４８：１０１

６－１０１９；オクセンバーグ（Ｏｋｓｅｎｂｅｒｇ）ら、１９９０、Ｎａｔｕｒｅ、３４５：３４４－３４６；を見ること。

【００８３】

強皮症において、その病気に関わるＴＡＲｓのいくつかの特定可変領域は特徴付けられていた。これらのＴＣＲｓはＶβ－６、Ｖβ－８、Ｖβ－１４、とＶα－１６、Ｖα－３Ｃ、Ｖα－７、Ｖα－１４、Ｖα－１５、Ｖα－１６、Ｖα－２８、Ｖα－１２を含む。従って、これらの蛋白質の少なくともひとつをコードしているＤＮＡ構築物のワクチン接種は、強皮症に関わるＴ細胞を標的とする免疫反応を誘発する。

【００８４】

Ｔ細胞媒介自己免疫疾患を患った患者、特にそのＴＣＲの可変領域がまだ特徴付けられていない患者を治療するために、滑膜組織診を行うことができる。Ｔ細胞のサンプルが使われ、Ｔ細胞媒介自己免疫疾患のＴＣＲｓのその可変領域は標準技術を用いて特定化された。遺伝子ウイルスはこの情報を用いて準備される。

【００８５】

Ｂ細胞は、全身性エリテマトーデス（ＳＬＥ）、バゼドウ病、重症筋無力病、自己免疫溶血性貧血、自己免疫血小板減少症、喘息、クリオグロブリン血症、原発性胆汁性硬化症、悪性貧血を含む自己免疫疾患を媒介した。これらの疾患は、内因性抗原と結合し、自己免疫疾患に関連した炎症性のカスケードを開始する抗体によって特徴付けられた。抗体の可変領域に対するワクチン接種は、その抗体を産生するＢ細胞を排除するためのＣＴＬｓを含む免疫反応を誘発する。

【００８６】

Ｂ細胞媒介自己免疫疾患を患った患者を治療するために、自己免疫活性化に関わるその抗体の可変領域は、特定化されるべきである。組織診が行われ、炎症部位での抗体のサンプルが利用され得る。Ｂ細胞媒介自己免疫疾患の抗体の可変領域は、標準技術を用いて特定化され得る。遺伝子ワクチンはこの情報を用いて準備され得る。

【００８７】

ＳＬＥの場合において、ひとつの抗原はＤＮＡであると信じられている。従って、ＳＬＥに対して免疫化された患者において、彼らの血清は抗ＤＮＡ抗体として選別され、血清内で見つけられたそのような抗ＤＮＡ抗体の可変領域をコードするＤＮＡ構築物を含むワクチンが準備され得る。

【００８８】

ＴＣＲｓと抗体の両方の可変領域における共通の構造特徴はよく知られている。特にＴＣＲもしくは抗体をコードしたそのＤＮＡ配列は、一般的に下記のよく知られた方法においてよく見られる。ここの参照に取り込まれている、カバット（Ｋａｂａｔ）ら、１９８７、Ｓｅｑｕｅｎｃｅ ｏｆ Ｐｒｏｔｅｉｎｓ ｏｆ Ｉｍｍｕｎｏｌｏｇｉｃａｌ Ｉｎｔｅｒｅｓｔ、Ｕ．Ｓ．Ｄｅｐａｒｔｍｅｎｔ ｏｆ Ｈｅａｌｔｈ ａｎｄ Ｈｕｍａｎ Ｓｅｒｖｉｃｅｓ、ＢａｔｈｅｓｄａＭＤ、に記載されている。付け加えると、ここの参照に取り込まれている、抗体からの機能的可変領域のクローニングの一般的な方法は、Ｃｈａｕｄｈａｒｙら、１９９０、Ｐｒｏｃ．Ｎａｔｌ．Ａｃａｄ．Ｓｃｉ．ＵＳＡ、８７：１０６６、内に見られる。

【００８９】

遺伝子療法に関連した発明のいくつかの実施例において、遺伝子構築物は代償遺伝子だけでなく、治療性蛋白質をコードする遺伝子を含んでいる。代償遺伝子の例としては、ジストロフィンまたは機能的断片をコードした遺伝子、嚢胞性線維症を患った患者における欠損遺伝子を代償するための遺伝子、インスリン遺伝子、ＡＤＡを患った患者における欠損遺伝子を代償するための遺伝子、ＦａｃｔｏｒＶＩＩＩをコードした遺伝子を含む。治療性蛋白質をコードした遺伝子の例として、エリスロポエチン、インターフェロン、ＬＤＬ受容体、ＧＭ－ＣＳＦ、ＩＬ－２、ＩＬ－４、ＴＮＦをコードした遺伝子を含む。その上、毒性物質に特に結合する一本鎖抗体成分をコードする遺伝的構築物は投与され得る。いくつかの好ましい実施例において、ジストロフィン遺伝子はミニ遺伝子の一部分として供

給され、筋ジストロフィーを患った個人の治療に用いられる。いくつかの好ましい実施例において、部分的ジストロフィン蛋白質に対するコード配列を含むミニ遺伝子が供給される。ジストロフィン異常は軽症ベッカー型筋ジストロフィー（ＢＭＤ）と重症デュシェンヌ型筋ジストロフィー（ＤＭＤ）の両方の原因である。ＢＭＤにおいてジストロフィンは作られるが、サイズ及びもしくは量ともに異常である。その患者は軽症ないし中程度に筋力が弱っている。ＤＭＤにおいてはジストロフィン蛋白質は作られず、患者は１３歳までにｃｈａｉｒ－ｂｏｕｎｄ（寝たきり、ベットで体を起こせる状態）になり、通常２０歳までに死んでしまう。何人かの患者において、特にＢＭＢを患っている人はにおいて、この発明によってもたらされたミニ遺伝子の発現によって産生される部分的ジストロフィン蛋白質は、改良された筋機能を供給できる。

【００９０】

いくつかの好ましい実施例において、ＩＬ－２、ＩＬ－４、インターフェロン、ＴＮＦをコードする遺伝子は、存在するだけでなく除去された腫瘍細胞に運ばれ、個人に再導入される。いくつかの好ましい実施例において、ガンマインターフェロンをコードする遺伝子は多発性硬化症を患った個人に投与される。

【００９１】

遺伝的ワクチンを改善するための修飾型核酸配列の使い方に加えて、この発明は弱毒生ワクチンの改善と、抗原をコードする外来性遺伝子を運ぶための組換えベクターを用いたワクチンの改善に関連している。弱毒生ワクチンと、外来性抗原を運ぶための組換えベクターの例としては、ここの参照に取り込まれている、米国特許番号：４，７２２，８４８；５，０１７，４８７；５，０７７，０４４；５，１１０，５８７；５，１１２，７４９；５，１７４，９９３；５，２２３，４２４；５，２２５，３３６；５，２４０，７０３；５，２４２，８２９；５，２９４，４４１；５，２９４，５４８；５，３１０，６６８；５，３１０，６６８；５，３８７，７４４；５，３８９，３６８；５，４２４，０６５；５，４５１，４９９；５，４５３，３６４；５，４６２，７３４；４，４７０，７３４；５，４８２，７１３に記載されている。提供される遺伝子構築物は、ワクチン内で発現を起こすために機能し得る調節配列と連結可能な修飾型ヌクレオチド配列を含む。その遺伝子構築物は弱毒生ワクチンと、この発明による改良されたワクチンを産生するための組換えワクチンへ取り込まれる。同様に修飾型核酸配列は要求された蛋白質をコードする遺伝子治療学を行うのに有用な組換えベクターにおいて使われ得る。

【００９２】

この発明は、ＤＮＡワクチン、弱毒生ワクチン、組換えワクチンを含むワクチン成分の一部として、遺伝子構築物を個人の細胞へ運ぶ過程を有する個人を免疫化する改良された方法を供給する。この遺伝子構築物は免疫修飾蛋白質をコードするヌクレオチド配列と、ワクチン内において発現を起こすために機能し得る調節配列と連結可能なヌクレオチド配列を有する。この改善されたワクチンは細胞性免疫反応の増強に帰結する。

【００９３】

この発明は下記の例の方法によって説明され、発明のいくつかの実施例で詳しく述べるつもりである。これらの例は、その発明の範囲を制限するものとしての意図もなく、解釈されるものでもない。ここで特に記載された方法以外で本発明を実施することが可能であるということは明らかである。たくさんの修飾とこの発明の多様性は、ここでの指導の見解を可能にし、その結果、この発明の範囲内である。

【００９４】

【実施例】

例１：材料と方法

ｍＲＮＡ二次構造の予測

導入遺伝子の翻訳効率を上げるために、ＲＮＡ二次構造をＭｕｌＦｏｌｄを用いて予測し、マッキントッシュコンピューター用のＬｏｏｐＤｌｏｏｐソフトウェアを用いて見た。

【００９５】

放射線ラベル試験管内翻訳蛋白質の免疫沈降

^{35}Ｓラベル蛋白質産物はＴＮＴ－Ｔ７関連転写／翻訳系（Ｐｒｏｍｅｇａ）を用いて準備された。１０μｌの放射線ラベル蛋白質サンプルと１μｌ抗Ｈｉｓ（Ｃ－ｔｅｒｍ）抗体（Ｉｎｖｉｔｒｏｇｅｎ、ＣＡ）を３００μｌのＲＩＰＡバッファーに加え、静かに混ぜた。４℃、９０分インキュベーションした後、プロテインＡ－セファロースビーズ（Ａｍｅｒｓｈａｍ Ｐｈａｒｍａｃｉａ Ｂｉｏｔｅｃｈ、Ｐｉｓｃアデニンとチミンａｗａｙ、ＮＪ）を蛋白質－抗体複合体にチューブ当たり最終濃度の５ｍｇ加え、そのサンプルは回転性シェーカーにおいて４℃、９０分で反応させた。

【００９６】

そのビーズはＲＩＰＡバッファーで３回洗浄し、２ＸＳＤＳサンプルバッファーに混ぜた。免疫沈降された蛋白質複合体は、ＳＤＳ／ＰＡＧＥ（１５％）ゲルにおける短時間の煮沸と分離によってセファロースビーズから溶離された。蛋白質サンプルの移動度は市販の^{14}Ｃ－メチル化分子量マーカー（Ｓｉｇｍａ－Ａｌｄｒｉｃｈ Ｃｏｒｐ．、Ｓｔ．Ｌｏｕｉｓ、ＭＯ）の移動度と比較された。そのゲルは固定され、１Ｍサリチル酸ナトリウム溶液で簡単に処理され、ゲルドライアー（ＢｉｏＲａｄ、Ｈｅｒｃｕｌｅｓ、ＣＡ）で乾燥された。その乾燥したゲルはＸ－線フィルム（Ｋｏｄａｋ、Ｒｏｃｈｅｓｔｅｒ、ＮＹ）に一晩露光された。試験管内翻訳蛋白質の分子サイズは２１．５ｋＤであった。

【００９７】

試験管内翻訳蛋白質

非放射性試験管内翻訳Ｃｐ蛋白質は非放射性構築物を持つＴＮＴ－Ｔ７関連転写／翻訳系（Ｐｒｏｍｅｇａ、Ｍａｄｉｓｏｎ、ＷＩ）を用いて上記のように作成された。試験管内翻訳の対照には発現可能な挿入断片を欠いているｐｃＤＮＡ３．１ベクター（Ｉｎｖｉｔｒｏｇｅｎ、Ｓａｎ Ｄｉｅｇｏ、ＣＡ）の入った試験管内翻訳キットを用いて作成された。

【００９８】

マウスのＤＮＡ接種

６－から８－週目の雌ＢＡＬＢ／ｃマウス（Ｈａｒｌａｎ Ｓｐｒａｇｕｅ Ｄａｗｌｅｙ、Ｉｎｃ、．Ｉｎｄｉａｎａｐｏｌｉｓ、ＩＮ）の四頭筋にリン酸緩衝食塩水（ＰＢＳ）と０．２５％ブピバカイン－ＨＣｌ（Ｓｉｇｍａ、Ｓｔ．Ｌｏｕｉｓ、ＭＯ）に混ぜた１００μｇのｐＷＮＶｈ－ＤＪＹ、ｐＷＮＶｙ－ＤＪＹ、もしくはｐｃＤＮＡ３．１を注射した。マウスは２つのＤＮＡ免疫化（それぞれ１００μｇ）を２週間に分けて注射された。追加免疫後１３日目に、マウスは屠殺され、脾臓を回収され、そのリンパ球は分離、細胞免疫反応によって検定された。

【００９９】

フローサイトメトリーによる細胞内ＩＮＦ－γ検出

９６穴プレートのそれぞれのウェルに５０Ｕ／ｍｌ ｒＨｕＩＬ－２（Ｉｎｖｉｔｒｏｇｅｎ、Ｐｕｒｃｈａｓｅ、ＮＹ）、１０μｇ／ｍｌ Ｂｒｅｆｅｌｄｉｎ Ａ（Ｐｈａｒｍｉｎｇｅｎ、Ｓａｎ Ｄｉｅｇｏ、ＣＡ）、１００ｎｇ／ｍｌ ＰＭＡ（フォルボール１２－ミリスチン酸１３－酢酸）（Ｓｉｇｍａ、Ｓｔ．Ｌｏｕｉｓ、ＭＯ）、１μｇ／ｍｌ イオノマイシン（Ｓｉｇｍａ、Ｓｔ．Ｌｏｕｉｓ、ＭＯ）を含む１００μｌＲＰＭＩ－１６４０（５％ＦＢＳ添加）を入れた。試験管内翻訳蛋白質だけでなく試験管内翻訳対照（発現可能挿入断片を欠いたベクターを持つ試験管内翻訳キットを用いて作成された）４μｇ／ｍｌは５０μｌ Ｒ５培地に加えられた。抗原（Ａｇｓ）を添加後、分離された脾細胞はそれぞれのウェルに１Ｘ１０^{6}個、５０μｌ Ｒ５培地と共に添加された。フローサイトメトリーの補償のために、未処理マウスからの脾細胞はＩＬ－２とＢｒｅｆｅｌｄｉｎ Ａのみでセットアップされた。そのプレートは３７℃、５％CO_2インキュベーターにおいて５から６時間反応させた。対照として、細胞を抗原なしで反応させた。反応後、プレートは１２００ｒｐｍ、５分遠心され、上清は捨てた。細胞は２００μｌ ＰＢＳで懸濁、１％ ＢＳＡを添加、１５分氷上に放置、その後スピンダウンし、サンプル当たり０．１μｇの抗ＣＤ４－ＰＥｍＡｂ（Ｐｈａｒｍｉｎｇｅｎ）の溶けた５０μ ＰＢＳ／１％ＢＳＡで懸濁した。４℃、３０分で反応後、細胞はＰＢＳ／１％で２回洗浄した

。２回の洗浄後、細胞プレートは１００μｌのＣｙｔｏｆｉｘ／Ｃｙｔｏｐｅｒｍ溶液（Ｐｈａｒｍｉｎｇｅｎ）に懸濁され、４℃、２０分で反応した。細胞は１ＸＰｅｒｍ／Ｗａｓｈ（Ｐｈａｒｍｉｎｇｅｎ）で２回洗浄し、サンプル濃度当たり０。１μｇの抗ＩＮＦ－γ－ＡＰＣ（Ｐｈａｒｍｉｎｇｅｎ）を含む５０μｌ　Ｐｅｒｍ／Ｗａｓｈ溶液で懸濁した。４℃、３０分の反応後、細胞は１ＸＰｅｒｍ／Ｗａｓｈ溶液で２回洗浄し、２％パラホルムアルデヒドで固定、フローサイトメトリーで分析するまで４℃で保存した。

【０１００】

例２：ウエストナイルウイルスキャプシドｍＲＮＡへのリーダー配列の付加

ウエストナイルキャプシドｍＲＮＡにおいて自由エネルギーを最小化するためのリーダー配列の付加は、蛋白質発現と免疫反応の増強に帰結する。

【０１０１】

導入遺伝子の転写と翻訳効率を増強するために、ヒトＩｇＥリーダー配列はオープンリーディング枠（ｏｒｆ）配列の５’上流に付加された（Ｆｉｇ．１）。

【０１０２】

ヒト（ＷＮＶｙ－ＤＪＹ構築物（酵母コドン））においてそれほど優勢でなく使われているコドンを含むヒトＩｇＥリーダー配列をコードした配列の付加は、リーダー配列のないｍＲＮＡに対する二次構造（ＷＮＶｗｔ構築物（野生型））、もしくはヒトコドンで至適化されたリーダー配列をコードするｍＲＮＡに対する二次構造（ＷＮＶｈ－ＤＪＹ構築物（ヒトコドン））と比較して、自由エネルギー価の増加したｍＲＮＡに対する予測された二次構造に帰結する（Ｆｉｇ．２）。

【０１０３】

その上、放射線標識された試験管内翻訳蛋白質の免疫沈降によって決定された際、ヒト（酵母至適化）においてそれほど優勢でなく使われているコドンを含むリーダー配列をコードするその構築物は、ヒト至適化コドンを含むリーダー配列をコードする構築物よりもより高い蛋白質レベルをもたらす（Ｆｉｇ．３；Ｔａｂｌｅ１、酵母コドン利用）。酵母によってより優勢に用いられたコドンは一般的にアデニンとウラシルが豊富で、ヒトによってより優勢に用いられたコドンは一般的にグアニンとシトシンが豊富である（キムら、１９９７、Ｇｅｎｅ、ｓｕｐｒａ参照）。

【０１０４】

【表１】

表１　酵母コドンの優勢な使用

アミノ酸		酵母コドン
A	Ala	GCU
R	Arg	AGA
N	Asn	AAU
D	Asp	GAU
C	Cys	UGU
Q	Gln	CAA
E	Glu	GAA
G	Gly	GGU
H	His	CAU
I	Ile	AUU
L	Leu	UUA
K	Lys	AAA
P	Pro	CCA
F	Phe	UUU
S	Ser	UCU
T	Thr	ACU
W	Trp	UGG
Y	Tyr	UAU
V	Val	GUU

【０１０５】

細胞内ＩＦＮ－γフローサイトメトリー分析によって決定される際、マウス筋肉へのＤＮＡプラスミド挿入は抗原特異性、ＣＤ４$^{+}$ Ｔｈ細胞依存性免疫反応を誘導する。ＣＤ４$^{+}$ Ｔｈ細胞依存性、細胞内ＩＦＮ－γ産生はフローサイトメトリーによって定量化された。ｐＷＮＶｙ－ＤＪＹ（ｐＷＮＶＣｙ）免疫マウスから単離された脾細胞は、ｐＷＮＶｈ－ＤＪＹ９ｐＷＮＶＣｈ）免疫マウスから単離された脾細胞と比較して、試験管内翻訳Ｃｐ蛋白質での刺激におけるＩＦＮ－γが高いレベルで発現していた（Ｆｉｇ．４参照）。

【０１０６】

例３：最少予測自由エネルギーの増加によるＨＩＶ－１　ｐｏｌ　ＲＮＡにおけるＲＮＡ二次構造の除去

リーダーコード配列の付加と酵母至適化（ヒトにおいてそれほど優勢ではなく使われている）するためのコドンの変換の戦略は、ＨＩＶ－１　ｐｏｌ遺伝子に適応された。ヒトにおいてそれほど優勢でなく使われたコドンを持つＩｇＥリーダー配列をコードする核酸配列がＨＩＶ－１　ｐｏｌ遺伝子の５’末端に付加された時、最小化された転写エネルギーの予測自由エネルギーは増加した（Ｆｉｇ．５）。

【０１０７】

ＨＩＶ－１　ｐｏｌ構造において、ヌクレオチド（ｎｔ）１７３８とｎｔ１９３８の間に位置された安定二次構造のいくつかの領域はＭｕｌＦｏｌｄ分析によって予測された（Ｆｉｇ．６）。ｎｔ１７３８からｎｔ１９３８までの領域におけるコドンの、ヒトにおけるそれほど優勢ではなく使われているコドン（酵母至適化コドン）への変化は、その領域における二次構造の弱小化に帰結する。修飾型コドンを持つ領域に対する予測された二次構造は、オリジナル配列に対する予測された二次構造より高い自由エネルギーを持った（Ｆｉｇ．７）。付け加えて、ｐｏｌ遺伝子の最初の２００ヌクレオチドにおけるｍＲＮＡ二次構造の形成は、ヒトにおいて最も優勢に使われているコドン（ヒト至適化コドン）（ＨＩＶ－１　ｐｏｌ　ｈｕ）を含む転写産物と比較して、ヒトにおいてそれほど優勢ではなく使われているコドン（酵母至適化コドン）（ＨＩＶ－１　ｐｏｌ　ｙｔ）を用いて最小化された（Ｆｉｇ．８）。最小自由エネルギーは－５３．０ｋｃａｌから－２６．４ｋｃａｌへと劇的に増加した。

【０１０８】

例４：最小予測自由エネルギーの増加によるＨＩＶ－１　ｇａｇＲＮＡにおけるＲＮＡ二次構造の除去

安定二次構造のいくつかの領域はＨＩＶ－１　ｇａｇ構造遺伝子の転写産物に対するＭｕｌＦｏｌｄ分析によって予測され（Ｆｉｇ．９）、最小自由エネルギーはヒトにおいてそれほど優勢でなく使われたコドン（酵母至適化）を用いて増加した（－３５１．０７ｋｃａｌから－２８３．１１ｋｃａｌへ）（Ｆｉｇ．１０）。

【０１０９】

例５：最小予測自由エネルギーの増加によるＷＮＶ　ｅｎｖ　ＲＮＡにおけるＲＮＡ二次構造の除去

ウエストナイルウイルスエンベロープ（ｅｎｖ）遺伝子において、ヒトにおいてそれほど優勢でなく使われているコドン（酵母至適化、ＷＮＶｙｔ２００）を持つ遺伝子の最初の２００塩基対（ｂｐ）でのＲＮＡエネルギー最小化の戦略適応は、野生型ＷＮＶ　ｅｎｖ遺伝子（ＷＮＶｗｔ２００）の転写産物もしくは、ヒトにおいて最も優勢に使われているコドン（ＷＮＶｈｕ２００）で至適化された転写産物と比較して、同族転写産物の最小自由エネルギーが増加した（Ｆｉｇ．１１）。

【０１１０】

前述の例はこの発明を説明するつもりであり、この発明をどんな風にも限定するものではない。この技術におけるそれらのスキルはこの発明の精神と範囲内での修飾であると認識する。

【０１１１】

ここで引用されたすべての参照は、そのまま参照としてこれによって含まれている。

【図面の簡単な説明】

【図１】
ＲＮＡ二次構造に基づいて修飾型構造（ＷＮＶＣｈｕとＷＮＶＣｙ）を持つ、ウエストナイルウイルス（ＷＮＶ）　野生型キャプシド（Ｃｐ）蛋白質（ＷＮＶＣ）のヌクレオチドとアミノ酸配列を示している。分泌型ＩｇＥシグナルリーダー配列はＷＮＶＣ蛋白質と融合された。ＷＮＶＣのプロモーターと５’－近接領域の間のリニア配列で当然である様々な発現レベルを回避するために、ＷＮＣＶのアミノ酸２－６に対するリーダー配列とコドンは、酵母至適化コドン（ＭＮＶＣｙ）またはヒト至適化コドン（ＷＮＶＣｈｕ）によって修飾型。しかしながら、両方の構築物におけるＷＮＶキャプシド蛋白質に対するコード配列の残った部分は、ヒト至適化コドンによって修飾型。１）ｓＩｇＥリーダー配列をエンコードしている野生型ヌクレオチド配列（４．ｓＩｇＥｏｒｉ）（ＳＥＱ　ＩＤ　ＮＯ：１）、２）ｓＩｇＥリーダー配列のアミノ酸配列（ヌクレオチド配列は上に記載）（ＳＥＱ　ＩＤ　ＮＯ：２）、３）ＷＮＶキャプシド蛋白質に対するアミノ酸配列（最初のメチオニンを除く）（ＳＥＱ　ＩＤ　ＮＯ：３）、４）ＷＮＶＣｈｕ構造の配列をエンコードしているｓＩｇＥｈ－ＷＮＶキャプシド蛋白質のヌクレオチド配列（１．ｓＩｇＥｈ－ＷＮＶＣｈｕ）（ＳＥＱ　ＩＤ　ＮＯ：４）、が示されている。ＷＮＶＣｙ構造の中のｓＩｇＥｈ－ＷＮＶキャプシド蛋白質に対するコード配列（２．ｓＩｇＥｙ－ＷＮＶＣｙ）と野生型ＷＮＶキャプシドエンコード配列（３．ＷＮＶＣｗｔ）の違いは、ＷＮＶＣｈｕ構造のヌクレオチド配列の下に示される。
【図２】
１）ＷＶＮキャプシド蛋白質をエンコードする野生型ｍＲＮＡ（ＷＮＶｗｔ）、２）ヒト至適化コドン（ＷＮＶｙ－ＤＪＹ）を含むｓＩｇＥリーダー／ＷＮＶキャプシド蛋白質をエンコードするｍＲＮＡ、３）酵母至適化コドン（ＷＮＶｈ－ＤＪＹ）を含むｓＩｇＥリーダー／ＷＮＶキャプシド蛋白質をエンコードするｍＲＮＡ、の最初の７３ヌクレオチドに対する自由エネルギー価と共に、予測されるＲＮＡ二次構造をＭｕｌＦｏｌｄで示している。ＷＮＶｙ－ＤＪＹ配列によって明らかにされた最後のコドン（グリシンに対するＧＧＣ）は、ヒト至適化である。示されるように、”Ｔ”はＲＮＡ鎖において”Ｕ”に書き換えられる。ＷＮＶｈ－ＤＪＹとＷＮＶｙ－ＤＪＹにおける融合のｓＩｇＥリーダー部分をエンコードしたｍＲＮＡ鎖のヌクレオチドは、はっきりと示されている。
【図３】
異なるＷＮＶキァプシド蛋白質構造：ｐＷＮＶＣｈｕ（ＷＮＶＣｈｕとｐＷＮＶｈ－ＤＪＹとも呼ばれる）とｐＷＮＶＣｙｔ（ＷＮＶＣｙとｐＷＮＶｙ－ＤＪＹとも呼ばれる）、からの試験管内転写／翻訳産物を電気泳動で分離し、免疫沈降し、放射線標識したオートラジオグラフを示している。左側の最初のレーンは、分子量マーカーを含む。矢印はメインキャプシド蛋白質産物の位置を指し示している。Ｃ－末端ポリヒスチジンタグとの融合蛋白質は抗－Ｈｉｓ抗体を用いて免疫沈降された。
【図４】
ＤＮＡ免疫されたマウスからの試験管内刺激脾細胞における細胞内ＩＮＦ－γ発現のフローサイトメトリー解析を示している。示された値は、二重ポジティブ細胞のパーセンテージである。上のパネルにおいて、細胞はＩＮＦ－γとＣＤ４４で染色され；下のパネルにおいて、細胞はＣＤ４とＩＮＦ－γで染色された。トップのラベルは、以前マウスに免疫したベクター、プラス、脾細胞の試験管内再刺激に用いた刺激ベクターを示している。免疫するベクターは、ｐｃＤＮＡＩＩＩ（ｐｃＤＮＡ３．１）、ｐＷＮＶｈ－ＤＪＹ（ｐＷＮＶＣｈ）、ｐＷＮＶｙ－ＤＪＹ（ｐＷＮＶＣｙ）であった。”Ｎｏ　Ａｇ”は、脾細胞は試験管内翻訳対照（例２に描写している）と反応させた、ということを指し示し、”ｐｒｏｔｅｉｎ”は、脾細胞はｐＷＮＶｙ－ＤＪＹ発現構築物からの翻訳された試験管内Ｃｐ蛋白質産物と反応させた、ということを指し示している。
【図５】
ＨＩＶ－１　ｐｏｌ遺伝子（ｐｏｌｗｔ２００ｒｎ）に対する野生型ｍＲＮＡの最初の２００ヌクレオチドと、ヒトにおいてそれほど優勢でなく使われた（酵母至適化）（ｓＩｇｙ＋ｐｏｌｗｔ）コドンを持つＩｇＥリーダー配列をエンコードする５’配列を含むＨＩ

Ｖ－１　ｐｏｌ遺伝子に対するｍＲＮＡの最初の２００ヌクレオチド、に対するエネルギー最小限化に基づいた自由エネルギー価を持つ予測される二次構造をＭｕｌＦｏｌｄで示している。示されるように、”Ｔ”はＲＮＡ鎖において”Ｕ”に書き換えられる。

【図６】

ＨＩＶ－１　ｐｏｌ構造遺伝子に対するｍＲＮＡの予測される二次構造をＭｕｌＦｏｌｄで示してある。

【図７】

ヌクレオチド１７３８からヌクレオチド１９３８までの２００ヌクレオチド領域配列はヒトにおいてそれほど優勢でなく利用されたコドン（酵母至適化コドン）を持つために変えられた後のＨＩＶ－１　ｐｏｌ構造遺伝子に対するｍＲＮＡの予測される二次構造をＭｕｌＦｏｌｄで示している。

【図８】

予測される二次構造、ならびに、ヒト至適化コドン（ＨＩＶ－１　ｐｏｌ　ｈｕ）を含むＨＩＶ－１　ｐｏｌ遺伝子に対するｍＲＮＡの最初の２００ヌクレオチドと、ヒトにおいて利用されたコドン（酵母至適化コドン）（ＨＩＶ　Ｐｏｌ　ｙｔ）を含むＨＩＶ－１　ｐｏｌ遺伝子に対するｍＲＮＡにおける全体の自由エネルギー価のＭｕｌＦｏｌｄで示している。示されるように、”Ｔ”はＲＮＡ鎖において”Ｕ”に書き換えられる。

【図９】

予測される二次構造と、ＨＩＶ－１　ｇａｇ構造遺伝子に対するｍＲＮＡ転写における全体の自由エネルギー価をＭｕｌＦｏｌｄで示してある

【図１０】

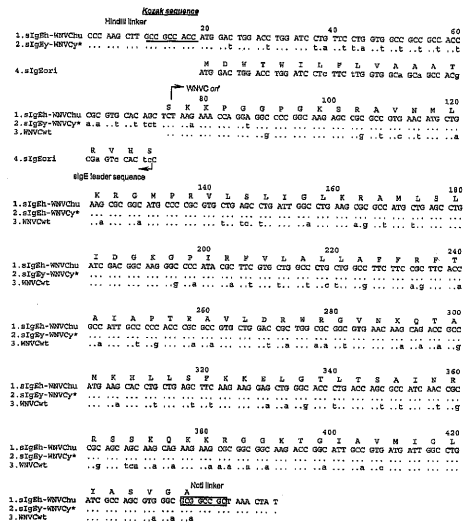
予測される二次構造と、ヒト（酵母至適化）においてそれほど優勢ではなく利用されたコドンに変えられたＨＩＶ－１　ｇａｇ構造遺伝子に対するｍＲＮＡ転写における全体の自由エネルギー価をＭｕｌＦｏｌｄで示してある。

【図１１】

予測される二次構造と、１）野生型ウエストナイルウイルス（ＷＮＶ）エンベロープ（ｅｎｖ）遺伝子（ＷＮＶｗｔ２００）、２）最も優勢に使われたヒトにおけるコドン（ＷＮＶｈｕ２００）に至適化されたＷＮＶ　ｅｎｖ遺伝子、３）ヒトにおいてそれほど優勢でなく利用されたコドン（酵母至適化、ＷＮＶｙｔ２００）を持つＷＮＶ　ｅｎｖ遺伝子、に対するｍＲＮＡ転写の最初の２００ヌクレオチドにおける全体の自由エネルギー価をＭｕｌＦｏｌｄで示してある。示されるように、”Ｔ”はＲＮＡ鎖において”Ｕ”に書き換えられる。

【図１】

Fig.1.



【図２】

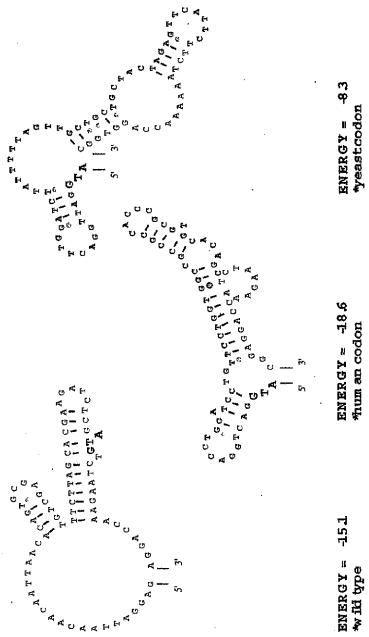
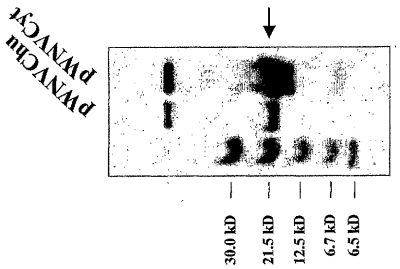


Fig. 2

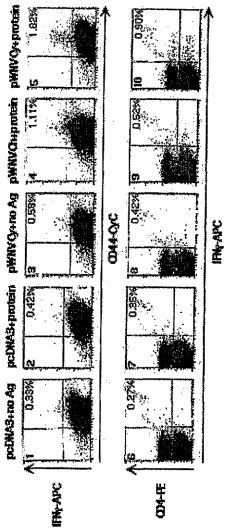
【図３】

Fig. 3 Immunoprecipitation of Radiolabeled *In Vitro* Translated Proteins

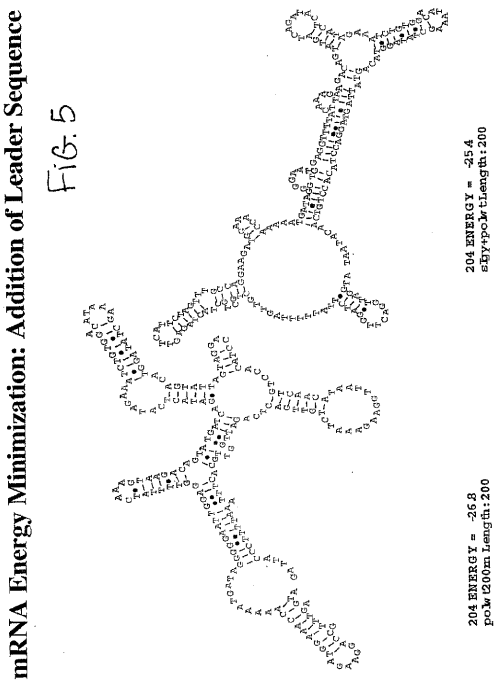


【図４】

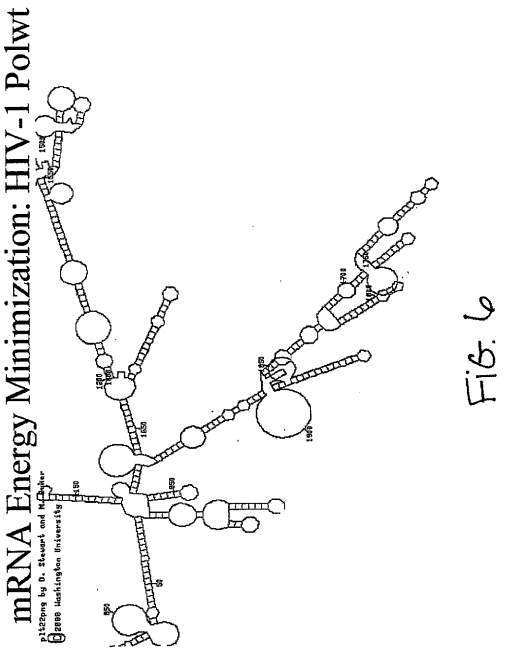
Fig. 4



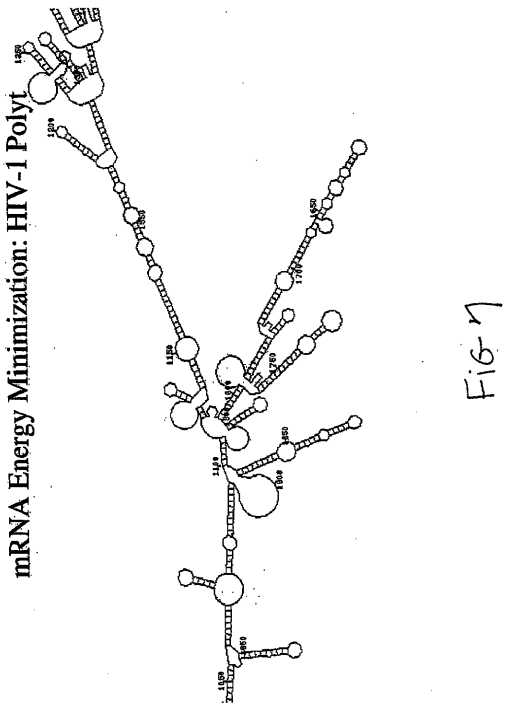
【図５】



【図６】

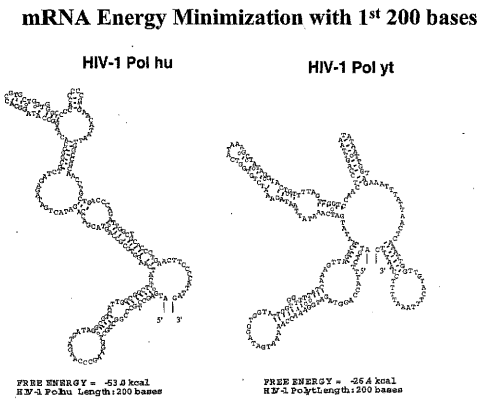


【図７】

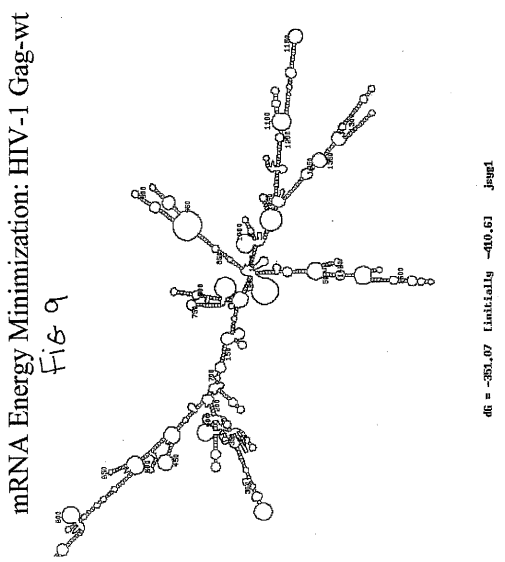


【図８】

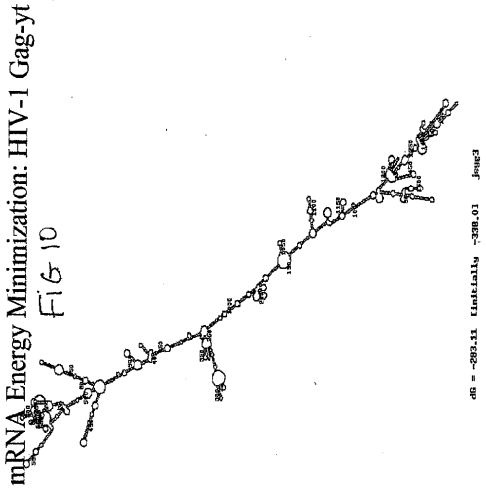
Fig 8



【図９】


mRNA Energy Minimization: HIV-1 Gag-wt
Fig 9

【図１０】


mRNA Energy Minimization: HIV-1 Gag-yt
Fig 10

【図１１】

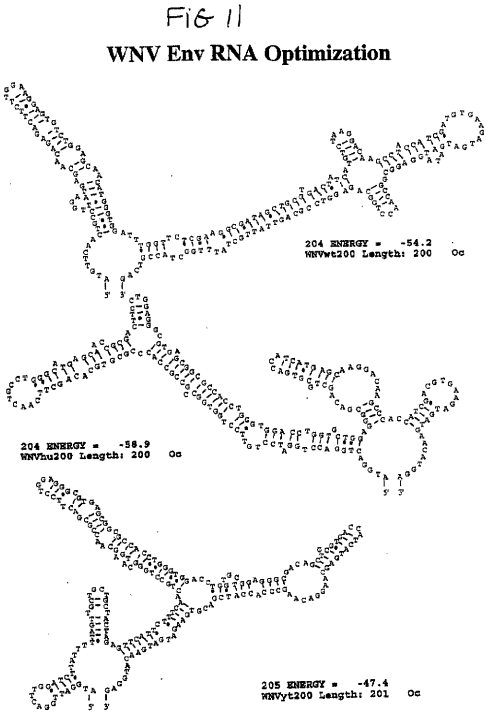


Fig 11
WNV Env RNA Optimization
204 ENERGY = -54.2
WNVwt200 Length: 200
204 ENERGY = -58.9
WNVhu200 Length: 200
205 ENERGY = -47.4
WNVyt200 Length: 201

【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
11 April 2002 (11.04.2002)

PCT

(10) International Publication Number
WO 02/29088 A2

(51) International Patent Classification[7]: C12Q

(21) International Application Number: PCT/US01/31451

(22) International Filing Date: 4 October 2001 (04.10.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/237,885 4 October 2000 (04.10.2000) US

(71) Applicant *(for all designated States except US)*: **THE TRUSTEES OF THE UNIVERSITY OF PENNSYLVANIA** [US/US]; Suite 300, 3700 Market Street, Philadelphia, PA 19104-3147 (US).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: **WEINER, David, B.** [US/US]; 717 Beacon Lane, Merion, PA 19066 (US). **YANG, Joo-Sung** [KR/US]; 102 South 42nd Street, Apartment 1, Philadelphia, PA 19104 (US).

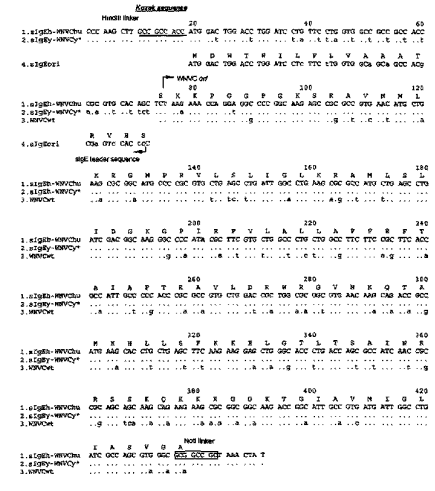
(74) Agents: **DeLUCA, Mark** et al.; Woodcock Washburn LLP, 46th floor, One Liberty Place, Philadelphia, PA 19103 (US).

(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, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian

[Continued on next page]

(54) Title: HIGHLY EXPRESSIBLE GENES



(57) Abstract: The present invention provides methods of producing protein in a recombinant expression system that comprises translation of mRNA transcribed from a heterologous DNA sequence in the expression system, said method comprising the steps of predicting the secondary structure of mRNA transcribed from a native heterologous DNA sequence wherein mRNA transcribed from the modified heterologous DNA sequence has a secondary structure having increased free energy compared to that of the secondary structure of the mRNA transcribed from the native heterologous DNA sequence; and using the modified heterologous DNA sequence in the recombinant expression system for protein production. The invention also provides injectable pharmaceutical compositions comprising a nucleic acid molecule that includes a modified coding sequence. The invention also provides coding sequence.

WO 02/29088 A2

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.

HIGHLY EXPRESSIBLE GENES

CROSS-REFERENCE TO RELATED APPLICATIONS

This application claims benefit under 35 U.S.C. §119(e) of U.S. Provisional Patent Application Serial No. 60/237,885, filed October 4, 2000, incorporated herein by reference.

FIELD OF THE INVENTION

The present invention relates generally to the fields of gene expression, gene therapy, and genetic immunization.

BACKGROUND OF THE INVENTION

The expression of a protein gene product is influenced by many factors, including gene copy number, gene integration site or gene location in the genome, transcription factors, mRNA stability, and translation efficiency. For example, the expression of the human immunodeficiency virus-1 (HIV-1) structural genes *gag*, *pol*, and *env* is dependent on the Rev/Rev-responsive element (RRE) at a posttranscriptional level. This dependency on Rev is a limiting factor for gene expression. In addition, highly stable RNA secondary structures that form in various regions of the HIV RNA transcript can block or otherwise interfere with ribosome movement, and thus effectively limit translation. Formation of stable RNA secondary structures in gene transcripts is a general phenomenon that can limit the translational yield of many protein gene products for a wide variety of genes.

Kim *et al.*, 1997, Gene, 199:293-301, which is incorporated herein by reference, optimized expression of human erythropoietin (EPO) in mammalian cells by altering the codons encoding the leader sequence and the first 6 amino acids of the mature EPO protein for the most

prevalently used yeast codons, and changing the codons encoding the rest of the EPO protein for the most prevalently used human codons.

U.S. patents US 5,972, 596 and 5,965,726 (Pavlakis *et al.*), which are incorporated herein by reference, describe methods of locating an inhibitory/instability sequence or sequences (INS: sequences that render an mRNA unstable or poorly utilized/translated) within the coding region of an mRNA and modifying the gene encoding the mRNA to remove the inhibitory/instability sequences with clustered nucleotide substitutions.

There is a need for new methods of expressing proteins and methods of increasing the level of protein expression of therapeutic and immunogenic transgenes. There is a need for methods of increasing the translational yields of any protein gene product. There is a need for methods of overcoming the limitations imposed by RNA secondary structure in RNA transcripts upon the ultimate level of protein expression of any gene. The present invention is directed to addressing these and other needs.

SUMMARY OF THE INVENTION

The present invention provides methods of producing protein in a recombinant expression system that comprises translation of mRNA transcribed from a heterologous DNA sequence in the expression system, said method comprising the steps of predicting the secondary structure of mRNA transcribed from a native heterologous DNA sequence; modifying the native heterologous DNA sequence to produce a modified heterologous DNA sequence wherein mRNA transcribed from the modified heterologous DNA sequence has a secondary structure having increased free energy compared to that of the secondary structure of the mRNA transcribed from the native heterologous DNA sequence; and using the modified heterologous DNA sequence in the recombinant expression system for protein production. The recombinant expression system may be a cell free *in vitro* transcription and translation system, an *in vitro* cell expression system, a DNA construct used in direct DNA injection, or a recombinant vector for delivery of DNA to an individual. The secondary structure of the mRNA transcribed from a native heterologous DNA sequence may be predicted using a computer and computer program. The native heterologous DNA sequence may be modified by increasing the AT content of the coding sequence, in particular, at the5' end of the coding sequence, or at the 5' end of the coding sequence within 200, 150, or 100 nucleotides from the initiation codon.

The present invention also provides injectable pharmaceutical compositions comprising a nucleic acid molecule that includes a modified coding sequence encoding a protein operably linked to regulatory elements, wherein the modified coding sequence comprises a higher AT or AU content relative to the AT or AU content of the native coding sequence, and further comprising a pharmaceutically acceptable carrier. The encoded proteins may be immunogens or non-immunogenic therapeutic proteins. The modifications may be within the first 100 to 200 bases of the coding sequence, within stretches of sequences dispersed throughout the coding sequence, or within in the last 100 to 200 bases.

The present invention also provides recombinant viral vectors comprising a nucleic acid molecule that includes a modified coding sequence encoding a protein operably linked to regulatory elements, wherein the modified coding sequence comprises a higher AT or AU content relative to the AT or AU content of the native coding sequence.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 presents the nucleotide and amino acid sequence of the West Nile Virus (WNV) wild type capsid (Cp) protein (WNVC) with constructs (WNVChu and WNVCy*) modified on the basis of RNA secondary structure. A secretory IgE signal leader sequence was fused to the WNVC protein. To avoid varied expression levels due to the linear sequence between the promoter and 5'-proximal region of the WNVC, the leader sequences and the codons for amino acids 2 - 6 of the WNVC were modified with yeast (WNVCy) or human (WNVChu) optimized codons. However, the remaining portion of the coding sequence for the WNV capsid protein, in both constructs, was modified with human optimized codons. Presented are 1) the wild type nucleotide sequence encoding the sIgE leader sequence (4. sIgEori) (SEQ ID NO:1), 2) the amino acid sequence of the sIgE leader sequence (appearing above the nucleotide sequence) (SEQ ID NO:2), 3) the amino acid sequence for the WNV capsid protein (minus the initial methionine) (SEQ ID NO:3), and 4) the nucleotide sequence of the sIgEh-WNV capsid protein encoding sequence of the WNVChu construct (1. sIgEh-WNVChu) (SEQ ID NO:4). Differences in the coding sequence for sIgEh-WNV capsid protein in the WNVCy construct (2. sIgEy-WNVCy*) and in the wild type WNV capsid encoding sequence (3. WNVCwt) are indicated below the nucleotide sequence of the WNVChu construct.

Figure 2 presents the MulFold predicted RNA secondary structures with free energy values for the first 73 nucleotides of 1) the wild type mRNA encoding WVN capsid protein

(WNVwt), 2) an mRNA encoding the sIgE leader/WNV capsid protein containing human optimized codons (WNVh-DJY), and 3) an mRNA encoding the sIgE leader/WNV capsid protein containing yeast optimized codons (WNVy-DJY). The last codon (GGC for glycine) shown for the WNVy-DJY sequence is human optimized. As shown, "T" represents "U" in the RNA strands. The nucleotides of the mRNA strands that encode the sIgE leader portion of the fusions in WNVh-DJY and WNVy-DJY are shown in bold.

Figure 3 presents an autoradiograph of electrophoretically separated, immunoprecipitated, radiolabeled *in vitro* transcription/translation products from two different WNV capsid protein constructs: pWNVChu (also called WNVChu and pWNVh-DJY) and pWNVCyt (also called WNVCy and pWNVy-DJY). The first lane on the left contains molecular weight markers. The arrow indicates the position of the main capsid protein product. The proteins, which are fusions with polyhistidine C-terminal tags, were immunoprecipitated using an anti-His antibody.

Figure 4 presents the flow cytometry analysis of intracellular IFN-γ expression in *in vitro* stimulated splenocytes from DNA immunized mice. Values presented are the percentage dual positive cells. In the upper panels, the cells were stained for INF-γ and CD44; in the lower panels the cells were stained for CD4 and IFN-γ. The labeling across the top indicates the vector used to immunize the mice plus the stimulus used for the *in vitro* restimulation of the splenocytes. The immunizing vectors were pcDNA3 (pcDNA3.1), pWNVh-DJY (pWNVCh), and pWNVy-DJY (pWNVCy). "No Ag" indicates that the splenocytes were incubated with an *in vitro* translation control (described in Example 2), "protein" indicates that the splenocytes were incubated with *in vitro* translated Cp protein product from the pWNVy-DJY expression construct.

Figure 5 presents the MulFold predicted RNA secondary structure with free energy values based upon energy minimization for the first 200 nucleotides of the wild type mRNA for to the HIV-1 *pol* gene (polwt200rn) and for the fist 200 nucleotides of an mRNA for HIV-1 *pol* gene including a 5' sequence encoding the IgE leader sequence with codons less prevalently used in humans (yeast optimized) (sIgy+polwt). As shown, "T" represents "U" in the RNA strand.

Figure 6 presents the MulFold predicted secondary structure of the mRNA for the HIV-1 *pol* structural gene.

Figure 7 presents the MulFold predicted secondary structure for the mRNA for the HIV-1 *pol* structural gene after the 200 nucleotide region of the sequence from nucleotide 1738 through

nucleotide 1938 has been altered to contain codons that are less prevalently utilized in humans (yeast optimized codons).

Figure 8 presents the MulFold predicted secondary structure and overall free energy value for the first 200 nucleotides of the mRNA for the HIV-1 *pol* gene containing human optimized codons (HIV-1 Pol hu), and for the mRNA for the HIV-1 *pol* gene containing codons less prevalently utilized in humans (yeast optimized codons) (HIV-1 Pol yt). As shown, "T" represents "U" in the RNA strands.

Figure 9 presents the MulFold predicted secondary structure and overall free energy value for the mRNA transcript for the HIV-1 *gag* structural gene.

Figure 10 presents the MulFold predicted secondary structure and overall free energy value for the mRNA transcript for the HIV-1 *gag* structural gene altered with codons that are utilized less prevalently in humans (yeast optimized).

Figure 11 presents the MulFold predicted secondary structures and overall free energy values for the first 200 nucleotides of the mRNA transcript for 1) the wild type West Nile Virus (WNV) envelope (*env*) gene (WNVwt200), 2) the WNV *env* gene optimized with the most prevalently used codons in humans (WNVhu200), and 3) the WNV *env* gene having codons that are utilized less prevalently in humans (yeast optimized, WNVyt200). As shown, "T" represents "U" in the RNA strands.

DESCRIPTION OF THE PREFERRED EMBODIMENTS

The present invention is based upon the discovery that enhancement of protein expression can be achieved by increasing the free energy of and destabilizing RNA secondary structure through changes at the nucleotide level. It has been discovered that an increase in the free energy (X kcal) of an RNA transcript will result in increased expression of the protein that it encodes. In preferred embodiments, an increase in the free energy (X kcal) within a 200 base segment of an RNA transcript will result in increased expression of the protein that it encodes. The segment is preferably at the 5' end, usually including the initiation codon. In some embodiments, the segment is preferably 200 bases, 150 bases, or 100 bases. The secondary structure of an RNA molecule is the collection of base pairs that occur in its three-dimensional structure. The secondary structure of a given RNA molecule can be predicted and such predicted secondary structure will have an assigned overall free energy value. It has been discovered that alterations to the primary sequence of an RNA transcript that result in an increase to the minimum predicted

overall free energy for a predicted secondary structure for that RNA, or that increase the minimum predicted free energy for a predicted secondary structure for regions of that RNA, will promote increased expression of the protein encoded by that RNA transcript. This strategy for the optimization of protein expression applies to any situation where expression is desired, including, but not limited to: *in vivo*, including, but not limited to, DNA vaccines, live vaccines, gene therapeutics, and transgenes; *in vitro*, including, but not limited to, recombinant manufacturing procedures using such systems as prokaryotic and eukaryotic (mammal, insect, and yeast) cells in culture; *ex vivo*, including, but not limited to, systems where cells receive expression constructs and are implanted into recipient organisms; and any other expression system where it is desirable to express a gene of interest or increase the expression of a gene.

One aspect of the invention is to generate an RNA encoding a protein that promotes efficient expression of that protein or that leads to increased levels of expression of the protein. Alterations to the sequence of the DNA encoding the RNA that lead to an increase in the minimum overall free energy for the predicted secondary structure of that RNA, or that increase the minimum free energy for the predicted secondary structure of one or more regions of that RNA promote efficient and/or increased expression of the encoded protein.

Increases to the free energy of the secondary structure of an RNA can be monitored by analyzing various altered versions of a sequence with a program like MFOLD, which calculates and predicts the most stable structure for an input sequence based upon energy minimization. MFOLD is computer software designed by Zuker, Jaeger, and colleagues (*see* Zuker, 1989, On finding all suboptimal foldings of an RNA molecule, Science, 244:48-52, and Jaeger *et al.*, 1989, Improved predictions of secondary structures for RNA, Proc. Natl. Acad. Sci. USA, 86:7706-7710, each of which is incorporated herein by reference) that is used for the prediction of RNA secondary structure by free energy minimization, using energy rules developed by Turner and colleagues (*see* Freier *et al.*, 1986, Proc. Natl. Acad. Sci. USA, 83:9373-9377, which is incorporated herein by reference). MulFold is the Macintosh version of MFOLD. LoopDloop is a secondary structure drawing program. The most stable structure will be the one with a minimum overall free energy. The more negative the value of the free energy for the structure, the more stable. Alterations to the sequence of the RNA that are predicted to result in a secondary structure having an overall higher free energy value, are destabilizing alterations which result in less stable RNA secondary structure and which promote efficient translation of the RNA and an increase in protein expression.

The practice of the present invention will employ, unless otherwise indicated, conventional methods of virology, immunology, microbiology, molecular biology, and recombinant DNA techniques within the skill of the art. Such techniques are explained fully in the literature. See, *e.g.*, Sambrook *et al.*, eds., Molecular Cloning: A Laboratory Manual (3rd ed.) Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY (2001); Ausubel *et al.*, eds., Current Protocols in Molecular Biology, John Wiley & Sons, New York, NY (2001); Glover & Hames, eds., DNA Cloning 3: A Practical Approach, Vols. I, II, & III, IRL Press, Oxford (1995); Colowick & Kaplan, eds., Methods in Enzymology, Academic Press; Weir *et al.*, eds., Handbook of Experimental Immunology, 5th ed., Blackwell Scientific Publications, Ltd., Edinburgh, (1997); Fields, Knipe, & Howley, eds., Fields Virology (3rd ed.) Vols. I & II, Lippincott Williams & Wilkins Pubs. (1996); Flint, *et al.*, eds., Principles of Virology: Molecular Biology, Pathogenesis, and Control, ASM Press, (1999); Coligan *et al.*, eds., Current Protocols in Immunology, John Wiley & Sons, New York, NY (2001), each of which is incorporated herein by reference.

Various definitions are made throughout this document. Most words have the meaning that would be attributed to those words by one skilled in the art. Words specifically defined either below or elsewhere in this document have the meaning provided in the context of the present invention as a whole and as typically understood by those skilled in the art.

As used herein, the term "recombinant expression system" refers to any nucleic acid based approach or system for the expression of a gene product or gene products of interest, that has been artificially organized (man made) of components directed toward the expression of the gene product or products. The components may be of naturally occurring genetic sources, synthetic or artificial, or some combination of natural and artificial genetic elements. Generally the gene product is a protein, polypeptide, or peptide. Examples of recombinant expression systems include, but are not limited to, a cell free *in vitro* transcription and translation system; an *in vitro* cell expression system; a DNA construct used in direct DNA injection; and a recombinant vector for delivery of DNA to an individual.

As used herein, the term "heterologous DNA sequences" refers to deoxyribonucleic acid based sequences that are in a non-natural context, for example, in a recombinant construct, plasmid, or virus, or inserted into a non-natural position in a chromosome, or introduced into a non-natural or foreign cell. "Heterologous DNA sequence" refers to any DNA sequence that is foreign or not naturally associated with the other DNA sequences to which it is associated or

linked (operably or otherwise), or a DNA sequence that is not naturally associated with the cell or organism into which it is introduced. An example of a heterologous DNA sequence is one that is used for the expression of a foreign or heterologous protein gene product in a host cell or organism. A heterologous DNA sequence can also be a part of a vector or expression construct having genetic material designed for directing the expression of a gene product, such as a protein, polypeptide, or peptide, in a host cell *in vivo* or *in vitro*, or in a cell free *in vitro* expression system.

As used herein, the term "native heterologous DNA sequence" refers to a heterologous DNA sequence that, although positioned in a non-natural context, has a nucleotide sequence that is not modified or altered from the sequence it has in its natural context. For example, a viral gene may be inserted into a recombinant expression construct, such that the viral gene is a heterologous DNA sequence with respect to other sequences in the construct, but without introduction of any changes to the nucleotide sequence of the viral gene. In this example, as a native heterologous DNA sequence, the viral gene has the native nucleotide sequence as would be found in its natural context within the genome of the virus, but the viral gene sequence is heterologous with respect to its new context. A native heterologous DNA sequence can also be any DNA sequence that is considered to be the reference or starting version of a DNA sequence, from which a modified (non-native) version of the DNA sequence, containing alterations to the nucleic acid sequence may be prepared. A native heterologous DNA sequence can also be composed of multiple native DNA sequences that are unaltered in sequence from that which is found in nature, but that are not naturally found together. An example of such a native heterologous DNA sequence composed of multiple native DNA sequences is a fusion gene composed of native genetic sequence from two different genes.

As used herein, the term "modified heterologous DNA sequence" refers to a heterologous DNA sequence that is not only positioned in a non-natural context, but also has a nucleotide sequence that is modified or altered from the sequence it has in its natural context. For example, a viral gene that is a modified heterologous DNA sequence will be inserted into a recombinant expression construct, such that the viral gene is heterologous with respect to other sequences in the construct, and further, will have a nucleotide sequence that is modified or altered and not the native nucleotide sequence as found in its natural context within the genome of the virus.

As used herein, the term "increased free energy" in reference to RNA secondary structure, refers to an increase in the free energy value for an RNA secondary structure. Free energy values that are more negative are lower than values that are less negative.

As used herein, the term "modified coding sequence" refers to a nucleic acid sequence (DNA- or RNA-based), that encodes a gene product, protein, polypeptide, or peptide, and that has been modified or altered from the native or naturally-occurring coding sequence for that gene product, protein, polypeptide, or peptide. The coding sequence may be comprised of sequences from more than one genetic source, for example, the coding sequence may be a fusion gene encoding a fusion protein having a leader sequence from a gene for one protein and the remaining sequence from a gene for another protein, brought together as one hybrid coding sequence, that is non-natural. In the case of such an example of a coding sequence comprised of sequences from more than one genetic source, "modified coding sequence" indicates that any modification is relative to the native or naturally-occurring coding sequence for the respective separate sequences.

As used herein, the term "native coding sequence" refers to a nucleic acid sequence (DNA- or RNA-based), that encodes a gene product, protein, polypeptide, or peptide, and that has not been modified or altered from the native or naturally-occurring coding sequence for that gene product, protein, polypeptide, or peptide. If the coding sequence encodes a fusion protein, the component parts have not been modified or altered from the native or naturally-occurring coding sequence for those component parts.

As used herein, the term "higher AT or AU content" refers to modifications to a coding sequence which render it a "modified coding sequence" such that if it is DNA-based it has a higher concentration of adenine and thymidine residues than the corresponding native coding sequence, and if it is RNA-based it has a higher concentration of adenine and uridine residues than the corresponding native coding sequence.

As used herein, the term "the first 200 bases" in reference to a modified coding sequence that has been modified relative to the native coding sequence, refers to the first 200 contiguous nucleotide bases from the 5' end of the respective coding sequence.

As used herein, the term "the first 150 bases" in reference to a modified coding sequence that has been modified relative to the native coding sequence, refers to the first 150 contiguous nucleotide bases from the 5' end of the respective coding sequence.

As used herein, the term "the first 100 bases" in reference to a modified coding sequence that has been modified relative to the native coding sequence, refers to the first 100 contiguous nucleotide bases from the 5' end of the respective coding sequence.

As used herein, the term "the last 200 bases" in reference to a modified coding sequence that has been modified relative to the native coding sequence, refers to the last 200 contiguous nucleotide bases from the 3' end of the respective coding sequence.

As used herein, the term "the last 150 bases" in reference to a modified coding sequence that has been modified relative to the native coding sequence, refers to the last 150 contiguous nucleotide bases from the 3' end of the respective coding sequence.

As used herein, the term "the last 100 bases" in reference to a modified coding sequence that has been modified relative to the native coding sequence, refers to the last 100 contiguous nucleotide bases from the 3' end of the respective coding sequence.

As used herein, the term "region of up to 200 bases in length" in reference to a coding sequence, refers to a region of up to 200 contiguous nucleotide bases of the coding sequence. The region may be anywhere within the coding sequence.

As used herein, the term "region of up to 150 bases in length" in reference to a coding sequence, refers to a region of up to 150 contiguous nucleotide bases of the coding sequence. The region may be anywhere within the coding sequence.

As used herein, the term "region of up to 100 bases in length" in reference to a coding sequence, refers to a region of up to 100 contiguous nucleotide bases of the coding sequence. The region may be anywhere within the coding sequence.

As used herein, the term "dispersed modifications" refers to any combination of at least two regions of contiguous nucleotide bases that are modified to have a higher AT or AU content relative to the native coding sequence in the respective regions, and that are dispersed throughout the sequence such that regions of modified coding sequences will alternate with regions of native coding sequence. By way of non-limiting example, a modified coding sequence may contain alternating regions of modifications, wherein the first 200 contiguous bases of the coding sequence have a higher AT or AU content relative to the native coding sequence, the next 200 bases of the coding sequence are non-modified relative to the native coding sequence, and the subsequent 200 contiguous base region is modified to have a higher AT or AU content relative to the native coding sequence. The size of the modified regions may be of any length, and is preferably 200, 150, or 100 bases in length. The size of non-modified regions will be of variable

length depending on the positioning of the modified regions. In preferred embodiments the dispersed modifications comprise alternating regions of modified and native coding sequence over the entire coding sequence, where the size of each alternating region is preferably 200 or 150 or 100 bases in length.

As used herein, "injectable pharmaceutical composition" refers to pharmaceutically acceptable compositions for use in patients that are sterile, pyrogen-free, and essentially free of any particulates or particulate matter. *See, Remington's Pharmaceutical Sciences*, 18th Ed., Gennaro, ed., Mack Publishing Co., Easton, PA, 1990 and U.S.P., the standards of the U. S. Pharmacopeia, which is incorporated herein by reference.

As used herein, "pharmaceutically acceptable carrier" includes any carrier that does not itself induce a harmful effect to the individual receiving the composition. For example, a "pharmaceutically acceptable carrier" should not induce the production of antibodies harmful to the recipient. Suitable "pharmaceutically acceptable carriers" are known to those of skill in the art and are described in *Remington's Pharmaceutical Sciences, supra.*

As used herein the term "target protein" is meant to refer to peptides and proteins encoded by gene constructs of the present invention which act as target proteins for an immune response. The terms "target protein" and "immunogen" are used interchangeably and refer to a protein against which an immune response can be elicited. The target protein is an immunogenic protein which shares at least an epitope with a protein from the pathogen or undesirable cell-type such as a cancer cell or a cell involved in autoimmune disease against which an immune response is desired. The immune response directed against the target protein will protect the individual against and/or treat the individual for the specific infection or disease with which the target protein is associated.

As used herein the term "desired protein" is meant to refer to peptides and proteins encoded by gene constructs of the present invention which either act as target proteins for an immune response or as a therapeutic or compensating protein in gene therapy regimens.

As used herein, the phrase "immunogenic fragment thereof" in reference to an immunogen, refers to fragments of less than the full length of the immunogen against which an immune response can be induced.

As used herein, the term "cancer antigens" refers to any proteins, polypeptides, or peptides, and the like, that are associated with and/or serve as markers for cancer, tumors, or cancer cells.

As used herein, the term "autoimmune disease associated proteins" refers to any proteins, polypeptides, or peptides, and the like, that are associated with and/or serve as markers for cells involved in and/or responsible for an autoimmune disease.

As used herein, the term "non-immunogenic therapeutic protein" refers to such proteins, polypeptides, and peptides that are useful for therapeutic treatment of various diseases and disorders, and to which an immune response is not desired and/or not expected upon their introduction into the body of a recipient organism, patient, or individual in need of such therapy or treatment. Examples of "non-immunogenic therapeutic proteins" are proteins that are missing or in low concentration in an individual having a genetic defect in the endogenous gene encoding the protein. Examples of "non-immunogenic therapeutic proteins" include, but are not limited to, cytokines, growth factors, blood products, and enzymes.

As used herein, the term "recombinant viral vector" refers to a construct, based upon the genome of a virus, that can be used as a vehicle for the delivery of nucleic acids encoding proteins, polypeptides, or peptides of interest. Recombinant viral vectors are well known in the art and are widely reported. Recombinant viral vectors include, but are not limited to, retroviral vectors, adenovirus vectors, and adeno-associated virus vectors, which are prepared using routine methods and starting materials.

As used herein, the term "genetic construct" refers to the DNA or RNA molecules that comprise a nucleotide sequence which encodes a target protein or immunomodulating protein. The coding sequence includes initiation and termination signals operably linked to regulatory elements including a promoter and polyadenylation signal capable of directing expression in the cells of the individual to whom the nucleic acid molecule is administered.

As used herein, the term "expressible form" refers to gene constructs which contain the necessary regulatory elements operably linked to a coding sequence that encodes a target protein or an immunomodulating protein, such that when present in the cell of the individual, the coding sequence will be expressed.

As used herein, the term "sharing an epitope" refers to proteins which comprise at least one epitope that is identical to or substantially similar to an epitope of another protein.

As used herein, the term "substantially similar epitope" is meant to refer to an epitope that has a structure which is not identical to an epitope of a protein but nonetheless invokes a cellular or humoral immune response which cross reacts to that protein.

As used herein, the term "intracellular pathogen" is meant to refer to a virus or pathogenic organism that, during at least part of its reproductive or life cycle, exists within a host cell and therein produces or causes to be produced, pathogen proteins.

As used herein, the term "hyperproliferative diseases" is meant to refer to those diseases and disorders characterized by hyperproliferation of cells.

As used herein, the term "hyperproliferative-associated protein" is meant to refer to proteins that are associated with a hyperproliferative disease and/or hyperproliferative cells.

In some preferred embodiments, it is preferred that the alterations to the RNA do not alter the sequence of the protein. In some preferred embodiments, it is preferred that the 200 bases, within which the alterations are introduced, are at the 5' end of the RNA transcript. In some embodiments, it is preferred to increase the free energy in more than one segment of the RNA transcript. Optionally, a leader sequence may be added to increase the free energy of the secondary structure of the RNA.

A stable RNA secondary structure at the 5' end of open reading frame (orf) sequences may block efficient transcription by interfering with ribosome function. Many RNAs have highly stable secondary structural integrity, and these interactions can inhibit gene expression Addition of a sequence encoding a leader, modified such that it was optimized with an AT-rich sequence, resulted in a higher free energy for the predicted RNA structure and allowed efficient initiation by the cellular ribosomes. The stable RNA secondary structure is removed by increasing the free energy.

Therefore, according to the present invention, increasing the AU content in a coding sequence optimizes the sequence by reducing the corresponding RNA secondary structure's integrity, and resulting in increased protein expression/translation, by melting of the inhibitory secondary structures (stem loops) in the RNA transcript. The disruption of secondary structure integrity is particularly important in the 5' portion of the RNA or coding sequence, particularly the first 100 to 200 nucleotides of the RNA. In some embodiments, the AU or AT content is increased in the first 100 to 200 nucletides from the initiation of transcription, and in some embodiments the AU or AT content is increased in the first 100 to 200 nucleotides of the coding sequence or start of translation. In some embodiments, the disruption of secondary structure integrity of the RNA is achieved by full gene changes or alternating patterns within 100 to 200 nucleotide base stretches. Modification of the 3' end is also important.

The strategy of adding a leader-encoding sequence and altering the codons of that sequence to be yeast optimized (less frequently used codons in humans) is applicable to any gene encoding any protein, for example genes encoding viral proteins, including, but not limited to, the HIV-1 *pol* gene. AU-rich content is preferred; human dominant codons/high GC content is not preferred. It has been discovered that lowering the stability of regions of secondary structure within mRNAs can be accomplished without prior knowledge of protein expression or structure. The resultant increased minimum free energy of the secondary structure that is predicted to form from the altered transcript renders the altered transcript capable of enhanced protein expression over the original.

Using standard techniques and readily available starting materials, a modified nucleic acid molecule may be prepared. The nucleic acid molecule may be incorporated into an expression vector which is then incorporated into a host cell. Host cells for use in well known recombinant expression systems for production of proteins are well known and readily available. Examples of host cells include bacteria cells such as *E. coli,* yeast cells such as *S. cerevisiae,* insect cells such as *S. frugiperda,* non-human mammalian tissue culture cells Chinese hamster ovary (CHO) cells and human tissue culture cells such as HeLa cells.

In some embodiments, for example, one having ordinary skill in the art can, using well known techniques, insert DNA molecules into a commercially available expression vector for use in well known expression systems. For example, the commercially available plasmid pSE420 (Invitrogen, San Diego, CA) may be used for production of immunomodulating proteins in *E. coli.* The commercially available plasmid pYES2 (Invitrogen, San Diego, CA) may, for example, be used for production in *S. cerevisiae* strains of yeast. The commercially available MAXBAC™ complete baculovirus expression system (Invitrogen, San Diego, CA) may, for example, be used for production in insect cells. The commercially available plasmid pcDNAI or pcDNA3 (Invitrogen, San Diego, CA) may, for example, be used for production in mammalian cells such as CHO cells. One having ordinary skill in the art can use these commercial expression vectors and systems or others to produce immunomodulating proteins by routine techniques and readily available starting materials. (*See e.g.*, Sambrook *et al.*, eds., 2001, *supra*) Thus, the desired proteins can be prepared in both prokaryotic and eukaryotic systems, resulting in a spectrum of processed forms of the protein.

One having ordinary skill in the art may use other commercially available expression vectors and systems or produce vectors using well known methods and readily available starting

materials. Expression systems containing the requisite control sequences, such as promoters and polyadenylation signals, and preferably enhancers, are readily available and known in the art for a variety of hosts (*See e.g.*, Sambrook *et al.*, eds., 2001, *supra*).

The expression vector including the modified DNA is used to transform the compatible host which is then cultured and maintained under conditions wherein expression of the foreign DNA takes place. The protein of the present invention thus produced is recovered from the culture, either by lysing the cells or from the culture medium as appropriate and known to those in the art. One having ordinary skill in the art can, using well known techniques, isolate the protein that is produced using such expression systems. The methods of purifying proteins from natural sources using antibodies may be equally applied to purifying protein produced by recombinant DNA methodology.

The pharmaceutical compositions of the present invention may be administered by any means that enables the active agent to reach the agent's site of action in the body of a mammal. The pharmaceutical compositions of the present invention may be administered in a number of ways depending upon whether local or systemic treatment is desired and upon the area to be treated. Administration may be topical (including ophthalmic, vaginal, rectal, intranasal, transdermal), oral or parenteral. Parenteral administration includes intravenous drip, subcutaneous, intraperitoneal or intramuscular injection, pulmonary administration, *e.g.*, by inhalation or insufflation, or intrathecal or intraventricular administration.

The present invention further relates to injectable pharmaceutical compositions which comprise such nucleic acid molecules.

The injectable pharmaceutical compositions that comprise a modified nucleotide sequence operably linked to regulatory elements may be delivered using any of several well known technologies including DNA injection (also referred to as DNA vaccination), recombinant vectors such as recombinant adenovirus, recombinant adenovirus associated virus and recombinant vaccinia.

DNA vaccines are described in U.S. Patent Nos. 5,593,972, 5,739,118, 5,817,637, 5,830,876, 5,962,428, 5,981,505, 5,580,859, 5,703,055, 5,676,594, and the priority applications cited therein, which are each incorporated herein by reference. In addition to the delivery protocols described in those applications, alternative methods of delivering DNA are described in U.S. Patent Nos. 4,945,050 and 5,036,006, which are both incorporated herein by reference.

Routes of administration include, but are not limited to, intramuscular, intranasally, intraperitoneal, intradermal, subcutaneous, intravenous, intraarterially, intraoccularly and oral as well as topically, transdermally, by inhalation or suppository or to mucosal tissue such as by lavage to vaginal, rectal, urethral, buccal and sublingual tissue. Preferred routes of administration include to mucosal tissue, intramuscular, intraperitoneal, intradermal and subcutaneous injection. Genetic constructs may be administered by means including, but not limited to, traditional syringes, needleless injection devices, or "microprojectile bombardment gene guns".

When taken up by a cell, the genetic construct(s) may remain present in the cell as a functioning extrachromosomal molecule and/or integrate into the cell's chromosomal DNA. DNA may be introduced into cells where it remains as separate genetic material in the form of a plasmid or plasmids. Alternatively, linear DNA which can integrate into the chromosome may be introduced into the cell. When introducing DNA into the cell, reagents which promote DNA integration into chromosomes may be added. DNA sequences which are useful to promote integration may also be included in the DNA molecule. Alternatively, RNA may be administered to the cell. It is also contemplated to provide the genetic construct as a linear minichromosome including a centromere, telomeres and an origin of replication. Gene constructs may remain part of the genetic material in attenuated live microorganisms or recombinant microbial vectors which live in cells. Gene constructs may be part of genomes of recombinant viral vaccines where the genetic material either integrates into the chromosome of the cell or remains extrachromosomal.

Genetic constructs include regulatory elements necessary for gene expression of a nucleic acid molecule. The elements include: a promoter, an initiation codon, a stop codon, and a polyadenylation signal. In addition, enhancers are often required for gene expression of the sequence that encodes the target protein. It is necessary that these elements be operable linked to the sequence that encodes the desired proteins and that the regulatory elements are operably in the individual to whom they are administered. Initiation codons and stop codon are generally considered to be part of a nucleotide sequence that encodes the desired protein. However, it is necessary that these elements are functional in the individual to whom the gene construct is administered. The initiation and termination codons must be in frame with the coding sequence. Promoters and polyadenylation signals used must be functional within the cells of the individual. Examples of promoters useful to practice the present invention, especially in the production of

a genetic vaccine for humans, include but are not limited to promoters from Simian Virus 40 (SV40), Mouse Mammary Tumor Virus (MMTV) promoter, Human Immunodeficiency Virus (HIV) such as the HIV Long Terminal Repeat (LTR) promoter, Moloney virus, ALV, Cytomegalovirus (CMV) such as the CMV immediate early promoter, Epstein Barr Virus (EBV), Rous Sarcoma Virus (RSV) as well as promoters from human genes such as human Actin, human Myosin, human Hemoglobin, human muscle creatine and human metalothionein. Examples of polyadenylation signals useful to practice the present invention, especially in the production of a genetic vaccine for humans, include but are not limited to SV40 polyadenylation signals and LTR polyadenylation signals. In particular, the SV40 polyadenylation signal which is in pCEP4 plasmid (Invitrogen, San Diego CA), referred to as the SV40 polyadenylation signal, is used. In addition to the regulatory elements required for DNA expression, other elements may also be included in the DNA molecule. Such additional elements include enhancers. The enhancer may be selected from the group including but not limited to: human Actin, human Myosin, human Hemoglobin, human muscle creatine and viral enhancers such as those from CMV, RSV and EBV. Genetic constructs can be provided with mammalian origin of replication in order to maintain the construct extrachromosomally and produce multiple copies of the construct in the cell. Plasmids pCEP4 and pREP4 from Invitrogen (San Diego, CA) contain the Epstein Barr virus origin of replication and nuclear antigen EBNA-1 coding region which produces high copy episomal replication without integration.

One method of the present invention comprises the steps of administering nucleic acid molecules intramuscularly, intranasally, intraperatoneally, subcutaneously, intradermally, or topically or by lavage to mucosal tissue selected from the group consisting of inhalation, vaginal, rectal, urethral, buccal and sublingual.

In some embodiments, the nucleic acid molecule is delivered to the cells in conjunction with administration of a polynucleotide function enhancer or a genetic vaccine facilitator agent. Polynucleotide function enhancers are described in U.S. Serial Number 08/008,342 filed January 26, 1993, U.S. Serial Number 08/029,336 filed March 11, 1993, U.S. Serial Number 08/125,012 filed September 21, 1993, and International Application Serial Number PCT/US94/00899 filed January 26, 1994, which are each incorporated herein by reference. Genetic vaccine facilitator (GVF) agents are described in U.S. Serial Number 08/221,579 filed April 1, 1994, which is incorporated herein by reference. The co-agents which are administered in conjunction with nucleic acid molecules may be administered as a mixture with the nucleic acid molecule or

administered separately simultaneously, before or after administration of nucleic acid molecules. In addition, other agents which may function as transfecting agents and/or replicating agents and/or inflammatory agents and which may be co-administered with a GVF include growth factors, cytokines and lymphokines such as α-interferon, gamma-interferon, platelet derived growth factor (PDGF), TNF, epidermal growth factor (EGF), IL-1, IL-2, IL-4, IL-6, IL-10 and IL-12 as well as fibroblast growth factor, surface active agents such as immune-stimulating complexes (ISCOMS), Freund's incomplete adjuvant, LPS analog including monophosphoryl Lipid A (MPL), muramyl peptides, quinone analogs and vesicles such as squalene and squalene, and hyaluronic acid may also be used administered in conjunction with the genetic construct. In some embodiments, an immunomodulating protein may be used as a GVF.

The pharmaceutical compositions according to the present invention comprise about 1 nanogram to about 2000 micrograms of DNA. In some preferred embodiments, pharmaceutical compositions according to the present invention comprise about 5 nanogram to about 1000 micrograms of DNA. In some preferred embodiments, the pharmaceutical compositions contain about 10 nanograms to about 800 micrograms of DNA. In some preferred embodiments, the pharmaceutical compositions contain about 0.1 to about 500 micrograms of DNA. In some preferred embodiments, the pharmaceutical compositions contain about 1 to about 350 micrograms of DNA. In some preferred embodiments, the pharmaceutical compositions contain about 25 to about 250 micrograms of DNA. In some preferred embodiments, the pharmaceutical compositions contain about 100 to about 200 micrograms DNA.

The pharmaceutical compositions according to the present invention are formulated according to the mode of administration to be used. In cases where pharmaceutical compositions are injectable pharmaceutical compositions, they are sterile, pyrogen free and particulate free. An isotonic formulation is preferably used. Generally, additives for isotonicity can include sodium chloride, dextrose, mannitol, sorbitol and lactose. In some cases, isotonic solutions such as phosphate buffered saline are preferred. Stabilizers include gelatin and albumin. In some embodiments, a vaso-constriction agent is added to the formulation.

The present invention is useful to elicit broad immune responses against a target protein, *i.e.*, proteins specifically associated with pathogens, allergens or the individual's own "abnormal" cells. The present invention is useful to immunize individuals against pathogenic agents and organisms such that an immune response against a pathogen protein provides protective immunity against the pathogen. The present invention is useful to combat hyperproliferative

diseases and disorders such as cancer by eliciting an immune response against a target protein that is specifically associated with the hyperproliferative cells. The present invention is useful to combat autoimmune diseases and disorders by eliciting an immune response against a target protein that is specifically associated with cells involved in the autoimmune condition.

The nucleic acid molecule(s) may be provided as plasmid DNA, the nucleic acid molecules of recombinant vectors or as part of the genetic material provided in an attenuated vaccine or cell vaccine. Alternatively, in some embodiments, the target protein and/or wither or both immunomodulating proteins may be delivered as a protein in addition to the nucleic acid molecules that encode them or instead of the nucleic acid molecules that encode them.

The present invention may be used to immunize an individual against all pathogens such as viruses, prokaryotic and pathogenic eukaryotic organisms such as unicellular pathogenic organisms and multicellular parasites. The present invention is particularly useful to immunize an individual against those pathogens which infect cells and which are not encapsulated such as viruses, and prokaryotes such as gonorrhea, listeria and shigella. In addition, the present invention is also useful to immunize an individual against protozoan pathogens which include a stage in the life cycle where they are intracellular pathogens.

In order to produce a genetic vaccine to protect against pathogen infection, genetic material which encodes immunogenic proteins against which a protective immune response can be mounted must be included in a genetic construct as the coding sequence for the target. Whether the pathogen infects intracellularly, for which the present invention is particularly useful, or extracellularly, it is unlikely that all pathogen antigens will elicit a protective response. Because DNA and RNA are both relatively small and can be produced relatively easily, the present invention provides the additional advantage of allowing for vaccination with multiple pathogen antigens. The genetic construct used in the genetic vaccine can include genetic material which encodes many pathogen antigens. For example, several viral genes may be included in a single construct thereby providing multiple targets.

Another aspect of the present invention provides a method of conferring a broad based protective immune response against hyperproliferating cells that are characteristic in hyperproliferative diseases and to a method of treating individuals suffering from hyperproliferative diseases. Examples of hyperproliferative diseases include all forms of cancer and psoriasis.

It has been discovered that introduction of a genetic construct that includes a nucleotide sequence which encodes an immunogenic "hyperproliferating cell"-associated protein into the cells of an individual results in the production of those proteins in the vaccinated cells of an individual. To immunize against hyperproliferative diseases, a genetic construct that includes a nucleotide sequence which encodes a protein that is associated with a hyperproliferative disease is administered to an individual.

In order for the hyperproliferative-associated protein to be an effective immunogenic target, it must be a protein that is produced exclusively or at higher levels in hyperproliferative cells as compared to normal cells. Target antigens include such proteins, fragments thereof and peptides which comprise at least an epitope found on such proteins. In some cases, a hyperproliferative-associated protein is the product of a mutation of a gene that encodes a protein. The mutated gene encodes a protein which is nearly identical to the normal protein except it has a slightly different amino acid sequence which results in a different epitope not found on the normal protein. Such target proteins include those which are proteins encoded by oncogenes such as *myb, myc, fyn,* and the translocation gene *bcr/abl, ras, src*, P53, *neu, trk* and EGRF. In addition to oncogene products as target antigens, target proteins for anti-cancer treatments and protective regimens include variable regions of antibodies made by B cell lymphomas and variable regions of T cell receptors of T cell lymphomas which, in some embodiments, are also used target antigens for autoimmune disease. Other tumor-associated proteins can be used as target proteins such as proteins which are found at higher levels in tumor cells including the protein recognized by monoclonal antibody 17-1A and folate binding proteins.

While the present invention may be used to immunize an individual against one or more of several forms of cancer, the present invention is particularly useful to prophylactically immunize an individual who is predisposed to develop a particular cancer or who has had cancer and is therefore susceptible to a relapse. Developments in genetics and technology as well as epidemiology allow for the determination of probability and risk assessment for the development of cancer in individual. Using genetic screening and/or family health histories, it is possible to predict the probability a particular individual has for developing any one of several types of cancer.

Similarly, those individuals who have already developed cancer and who have been treated to remove the cancer or are otherwise in remission are particularly susceptible to relapse

and reoccurrence. As part of a treatment regimen, such individuals can be immunized against the cancer that they have been diagnosed as having had in order to combat a recurrence. Thus, once it is known that an individual has had a type of cancer and is at risk of a relapse, they can be immunized in order to prepare their immune system to combat any future appearance of the cancer.

The present invention provides a method of treating individuals suffering from hyperproliferative diseases. In such methods, the introduction of genetic constructs serves as an immunotherapeutic, directing and promoting the immune system of the individual to combat hyperproliferative cells that produce the target protein.

The present invention provides a method of treating individuals suffering from autoimmune diseases and disorders by conferring a broad based protective immune response against targets that are associated with autoimmunity including cell receptors and cells which produce "self"-directed antibodies.

T cell mediated autoimmune diseases include rheumatoid arthritis (RA), multiple sclerosis (MS), Sjogren's syndrome, sarcoidosis, insulin dependent diabetes mellitus (IDDM), autoimmune thyroiditis, reactive arthritis, ankylosing spondylitis, scleroderma, polymyositis, dermatomyositis, psoriasis, vasculitis, Wegener's granulomatosis, Crohn's disease, and ulcerative colitis. Each of these diseases is characterized by T cell receptors that bind to endogenous antigens and initiate the inflammatory cascade associated with autoimmune diseases. Vaccination against the variable region of the T cells would elicit an immune response including CTLs to eliminate those T cells.

In RA, several specific variable regions of T cell receptors (TCRs) which are involved in the disease have been characterized. These TCRs include Vβ-3, Vβ-14, Vβ-17 and Vα-17. Thus, vaccination with a DNA construct that encodes at least one of these proteins will elicit an immune response that will target T cells involved in RA. *See*: Howell *et al.*, 1991, Proc. Natl. Acad. Sci. USA, 88:10921-10925; Paliard *et al.*, 1991, Science, 253:325-329; Williams *et al.*, 1992, J. Clin. Invest., 90:326-333; each of which is incorporated herein by reference.

In MS, several specific variable regions of TCRs which are involved in the disease have been characterized. These TCRs include Vβ-7 and Vα-10. Thus, vaccination with a DNA construct that encodes at least one of these proteins will elicit an immune response that will target T cells involved in MS. *See*: Wucherpfennig *et al.*, 1990, Science, 248:1016-1019; Oksenberg *et al.*, 1990, Nature, 345:344-346; each of which is incorporated herein by reference.

In scleroderma, several specific variable regions of TCRs which are involved in the disease have been characterized. These TCRs include Vβ-6, Vβ-8, Vβ-14 and Vα-16, Vα-3C, Vα-7, Vα-14, Vα-15, Vα-16, Vα-28 and Vα-12. Thus, vaccination with a DNA construct that encodes at least one of these proteins will elicit an immune response that will target T cells involved in scleroderma.

In order to treat patients suffering from a T cell mediated autoimmune disease, particularly those for which the variable region of the TCR has yet to be characterized, a synovial biopsy can be performed. Samples of the T cells present can be taken and the variable region of those TCRs identified using standard techniques. Genetic vaccines can be prepared using this information.

B cell mediated autoimmune diseases include systemic lupus erythematosus (SLE), Grave's disease, myasthenia gravis, autoimmune hemolytic anemia, autoimmune thrombocytopenia, asthma, cryoglobulinemia, primary biliary sclerosis, and pernicious anemia. Each of these diseases is characterized by antibodies which bind to endogenous antigens and initiate the inflammatory cascade associated with autoimmune diseases. Vaccination against the variable region of antibodies would elicit an immune response including CTLs to eliminate those B cells that produce the antibody.

In order to treat patients suffering from a B cell mediated autoimmune disease, the variable region of the antibodies involved in the autoimmune activity must be identified. A biopsy can be performed and samples of the antibodies present at a site of inflammation can be taken. The variable region of those antibodies can be identified using standard techniques. Genetic vaccines can be prepared using this information.

In the case of SLE, one antigen is believed to be DNA. Thus, in patients to be immunized against SLE, their sera can be screened for anti-DNA antibodies and a vaccine can be prepared which includes DNA constructs that encode the variable region of such anti-DNA antibodies found in the sera.

Common structural features among the variable regions of both TCRs and antibodies are well known. The DNA sequence encoding a particular TCR or antibody can generally be found following well known methods such as those described in Kabat *et al.*, 1987, *Sequence of Proteins of Immunological Interest*, U.S. Department of Health and Human Services, Bethesda MD, which is incorporated herein by reference. In addition, a general method for cloning

functional variable regions from antibodies can be found in Chaudhary *et al.*, 1990, Proc. Natl. Acad. Sci. USA, 87:1066, which is incorporated herein by reference.

In some of the embodiments of the invention that relate to gene therapy, the gene constructs contain either compensating genes or genes that encode therapeutic proteins. Examples of compensating genes include a gene which encodes dystrophin or a functional fragment, a gene to compensate for the defective gene in patients suffering from cystic fibrosis, an insulin, a gene to compensate for the defective gene in patients suffering from ADA, and a gene encoding Factor VIII. Examples of genes encoding therapeutic proteins include genes which encodes erythropoietin, interferon, LDL receptor, GM-CSF, IL-2, IL-4 and TNF. Additionally, genetic constructs which encode single chain antibody components which specifically bind to toxic substances can be administered. In some preferred embodiments, the dystrophin gene is provided as part of a mini-gene and used to treat individuals suffering from muscular dystrophy. In some preferred embodiments, a mini-gene which contains coding sequence for a partial dystrophin protein is provided. Dystrophin abnormalities are responsible for both the milder Becker's Muscular Dystrophy (BMD) and the severe Duchenne's Muscular Dystrophy (DMD). In BMD dystrophin is made, but it is abnormal in either size and/or amount. The patient is mild to moderately weak. In DMD no protein is made and the patient is chair-bound by age 13 and usually dies by age 20. In some patients, particularly those suffering from BMD, partial dystrophin protein produced by expression of a mini-gene delivered according to the present invention can provide improved muscle function.

In some preferred embodiments, genes encoding IL-2, IL-4, interferon, or TNF are delivered to tumor cells which are either present or removed and then reintroduced into an individual. In some embodiments, a gene encoding gamma interferon is administered to an individual suffering from multiple sclerosis.

In addition to using modified nucleic acid sequences to improve genetic vaccines, the present invention relates to improved attenuated live vaccines and improved vaccines which use recombinant vectors to deliver foreign genes that encode antigens. Examples of attenuated live vaccines and those using recombinant vectors to deliver foreign antigens are described in U.S. Patent Nos.: 4,722,848; 5,017,487; 5,077,044; 5,110,587; 5,112,749; 5,174,993; 5,223,424; 5,225,336; 5,240,703; 5,242,829; 5,294,441; 5,294,548; 5,310,668; 5,387,744; 5,389,368; 5,424,065; 5,451,499; 5,453,364; 5,462,734; 5,470,734; and 5,482,713, which are each incorporated herein by reference. Gene constructs are provided which include the modified

nucleotide sequence operably linked to regulatory sequences that can function in the vaccinee to effect expression. The gene constructs are incorporated in the attenuated live vaccines and recombinant vaccines to produce improved vaccines according to the invention. Likewise modified nucleic acid sequences can be used in recombinant vectors useful to deliver gene therapeutics that encode desired proteins.

The present invention provides an improved method of immunizing individuals that comprises the step of delivering gene constructs to the cells of individuals as part of vaccine compositions which include are provided which include DNA vaccines, attenuated live vaccines and recombinant vaccines. The gene constructs comprise a nucleotide sequence that encodes an immunomodulating protein and that is operably linked to regulatory sequences that can function in the vaccinee to effect expression. The improved vaccines result in an enhanced cellular immune response.

The invention is further illustrated by way of the following examples, which are intended to elaborate several embodiments of the invention. These examples are not intended, nor are they to be construed, as limiting the scope of the invention. It will be clear that the invention may be practiced otherwise than as particularly described herein. Numerous modifications and variations of the present invention are possible in view of the teachings herein and, therefore, are within the scope of the invention.

EXAMPLES

Example 1: Materials and Methods.

Prediction of mRNA secondary structure

To enhance translation efficiency of transgenes, RNA secondary structure was predicted by using MulFold and viewed by LoopDloop software for the Macintosh computer.

Immunoprecipitation of radiolabeled *in vitro* translated proteins

^{35}S-labeled protein products were prepared using the TNT-T7 coupled Transcription/Translation System (Promega). 10 ml of radiolabeled protein sample and 1 ml of anti-His (C-term) antibody (Invitrogen, CA) were added to 300 µl of RIPA buffer and mixed gently. After an incubation at 4°C for 90 minutes, Protein A-Sepharose beads (Amersham-Pharmacia Biotech, Piscataway, NJ) was added to the protein-antibody complexes at a final concentration of 5 mg per tube and the samples were then incubated at 4°C for 90 minutes in a rotating shaker. The beads were washed three times with RIPA buffer and suspended in 2X SDS

sample buffer. The immunoprecipitated protein complexes were eluted from the Sepharose beads by brief boiling and resolved in SDS/PAGE (15%) gels. The mobility of the protein samples was compared with that of commercially available ^{14}C-methylated molecular weight marker (Sigma-Aldrich Corp., St. Louis, MO). The gel was fixed, treated briefly with 1M sodium salicylate solution and dried in a gel drier (BioRad, Hercules, CA). The dried gel was exposed overnight to X-ray film (Kodak, Rochester, NY). The molecular size of the *in vitro* translated protein was 21.5kD.

***In vitro* translated protein**

Non-radioactive, *in vitro* translated Cp protein was also generated as described above, using the TNT-T7 coupled Transcription/Translation System (Promega, Madison, WI) with non-radioactive components. An *in vitro* translation control was generated using the *in vitro* translation kit with the pcDNA3.1 vector (Invitrogen, San Diego, CA), lacking an expressible insert.

DNA inoculation of mice

The quadriceps muscles of 6- to 8-week-old female BALB/c mice (Harlan Sprague Dawley, Inc., Indianapolis, IN) were injected with 100 μg of pWNVh-DJY, pWNVy-DJY, or pcDNA3.1 in phosphate buffered saline (PBS) and 0.25% bupivacaine-HCl (Sigma, St. Louis, MO). Mice were injected with two DNA immunizations (100 μg each) separated by two weeks. At thirteen days after the boost injection, the mice were sacrificed, the spleens were harvested, and the lymphocytes were isolated and tested for cellular immune responses.

Intracellular IFN-γ detection by flow cytometry

In each well of a round-bottom 96-well plate was placed 100 μl of RPMI-1640 (supplemented with 5 % FBS), containing 50 U/ml rHuIL-2 (Intergen, Purchase, NY), 10 μg/ml Brefeldin A (Pharmingen, San Diego, CA), 100 ng/ml PMA (Sigma, St. Louis, MO), and 1 μg/ml ionomycin (Sigma, St. Louis, MO). Either *in vitro* translated protein or an *in vitro* translation control (generated using the *in vitro* translation kit with the vector backbone lacking an expressible insert), at 4 μg/ml was added in 50 μl of R5 medium. After adding the antigens (Ags), isolated splenocytes were added to each well at $1x10^6$ cells in 50 μl of R5 medium. For the compensation in flow cytometry, splenocytes from naïve mice were set up with only IL-2 and Brefeldin A. The plates were incubated in 37°C, 5 % CO_2 in an incubator for 5 to 6 hours. As a control, cells were incubated without Ag. After incubation, the plate was spun at 1200 rpm for 5 minutes and the supernatants discarded. The cells were resuspended with 200 μl of PBS,

supplemented with 1% BSA, and put on ice for 15 minutes, and then spun down and resuspended with anti-CD4-PE mAb (Pharmingen) at 0.1 μg/sample in 50 μl of PBS/1% BSA. After incubation for 30 minutes at 4°C, the cells were washed twice with PBS/1%. After the second wash, cell pellets were resuspended with 100 μl of Cytofix/Cytoperm solution (Pharmingen) and incubated for 20 minutes at 4°C. The cells were washed twice with 1 x Perm/Wash (Pharmingen) and resuspended with 50 μl of Perm/Wash solution containing anti-IFN-γ-APC (Pharmingen) at 0.1 μg/sample concentration. After incubation for 30 minutes at 4°C, the cells were washed twice with 1x Perm/Wash solution and fixed with 2% paraformaldehyde, and then stored at 4°C until analyzed by flow cytometry.

Example 2: Addition of Leader Sequence to West Nile Virus Capsid mRNA.

The addition of a leader sequence to minimize free energy in the West Nile Virus Capsid mRNA resulted in enhanced protein expression and immune response.

To enhance the transcription and translation efficiency of transgenes, the human IgE leader sequence was added to the 5' upstream of open reading frame (orf) sequences (Fig.1).

The addition of a sequence encoding the human IgE leader sequence containing codons that are less prevalently utilized in humans (WNVy-DJY construct (yeast codon)) resulted in a predicted secondary structure for the mRNA having an increased free energy value, relative to the secondary structure for the mRNA without the leader sequence (WNVwt construct (wild type)), or relative to the secondary structure for the mRNA encoding a leader sequence optimized with human codons (WNVh-DJY construct (human codon)) (Fig. 2).

Furthermore, the construct encoding the leader sequence containing codons that are less prevalently utilized in humans (yeast optimized) yielded a higher level of protein than did the construct encoding the leader sequence containing human optimized codons, as determined by immunoprecipitation of radiolabeled *in vitro* translated proteins (Fig. 3; Table 1, yeast codon usage). The codons more prevalently used by yeast are, in general, AU rich; the codons more prevalently used by *Homo sapiens* are, in general, more GC rich (*see* Kim *et al.*, 1997, Gene, *supra*).

Table 1. Yeast codon prevalent usage.

Amino Acid		Yeast codon
A	Ala	GCU
R	Arg	AGA
N	Asn	AAU
D	Asp	GAU
C	Cys	UGU
Q	Gln	CAA
E	Glu	GAA
G	Gly	GGU
H	His	CAU
I	Ile	AUU
L	Leu	UUA
K	Lys	AAA
P	Pro	CCA
F	Phe	UUU
S	Ser	UCU
T	Thr	ACU
W	Trp	UGG
Y	Tyr	UAU
V	Val	GUU

DNA plasmid injection into mouse muscle induced an antigen-specific, $CD4^+$ Th cell-dependent immune response, as determined by intracellular IFN-γ/flow cytometry analysis. The $CD4^+$ Th cell-dependent, intracellular IFN-γ production was quantitated by flow cytometry. Splenocytes isolated from pWNVy-DJY (pWNVCy)-immunized mice, expressed higher levels of IFN-γ upon stimulation with *in vitro* translated Cp protein, than did the splenocytes isolated from pWNVh-DJY (pWNVCh)-immunized mice (see Fig. 4)

Example 3: Removal of RNA Secondary Structure in HIV-1 *pol* RNA by Increasing the Minimum Predicted Free Energy.

The strategy of adding a leader encoding sequence and altering the codons to be yeast optimized (less frequently used in human) was applied to the HIV-1 *pol* gene. When nucleic acid sequence encoding the IgE leader sequence with codons less prevalently used in humans (yeast optimized) was added to the 5' end of HIV-1 *pol* gene, the predicted free energy of the energy minimized transcript was increased (Fig. 5).

In HIV-1 *pol* structural gene, several regions of stable secondary structure, located between nucleotide (nt) 1738 and nt 1938, were predicted by MulFold analysis (Fig. 6). Alteration of the codons in the region from nt 1738 to nt 1938 to codons less prevalently utilized

in humans (yeast optimized codons) resulted in a weakening of the secondary structure in that region. The predicted secondary structure for the region with the modified codons had a higher free energy than the predicted secondary structure for the original sequence (Fig. 7). In addition, the formation of mRNA secondary structure in the first 200 nucleotides of the *pol* gene was minimized by using codons less prevalently utilized in humans (yeast optimized codons) (HIV-1 Pol yt), as compared to a transcript containing the most prevalently utilized codons in humans (human optimized codons) (HIV-1 Pol hu) (Fig. 8). The minimum free energy was dramatically increased from –53.0 kcal to –26.4 kcal.

Example 4: Removal of RNA Secondary Structure in *HIV-1 gag* RNA by Increasing the Minimum Predicted Free Energy.

Several regions of regions of stable secondary structure were predicted by MulFold analysis for the transcript for the HIV-1 *gag* structural gene (Fig. 9), and the minimum free energy was increased (from –351.07 kcal to –283.11 kcal) by using codons that are utilized less prevalently in humans (yeast optimized) (Fig. 10).

Example 5: Removal of RNA Secondary Structure in WNV *env* RNA by Increasing the Minimum Predicted Free Energy.

In the West Nile Virus envelope (*env*) gene, application of the strategy of mRNA energy minimization in the first 200 base pairs (bp) of the gene with codons that are utilized less prevalently in humans (yeast optimized, WNVyt200) increased the minimum free energy of the cognate transcript as compared to the transcript for the wild type WNV *env* gene (WNVwt200) or as compared to a transcript optimized with the most prevalently used codons in humans (WNVhu200) (Fig. 11).

The foregoing examples are meant to illustrate the invention and are not to be construed to limit the invention in any way. Those skilled in the art will recognize modifications that are within the spirit and scope of the invention.

All references cited herein are hereby incorporated by reference in their entirety.

What is claimed is:

1. A method of producing a protein in a recombinant expression system that comprises translation of mRNA transcribed from a heterologous DNA sequence in the expression system, said method comprising the steps of:

a) predicting the secondary structure of mRNA transcribed from a native heterologous DNA sequence;

b) modifying the native heterologous DNA sequence to produce a modified heterologous DNA sequence wherein mRNA transcribed from the modified heterologous DNA sequence has a secondary structure having increased free energy compared to that of the secondary structure of the mRNA transcribed from the native heterologous DNA sequence; and

c) using the modified heterologous DNA sequence in the recombinant expression system for protein production.

2. The method of claim 1, wherein the recombinant expression system is selected from the group consisting of: a cell free *in vitro* transcription and translation system; an *in vitro* cell expression system; a DNA construct used in direct DNA injection; and a recombinant vector for delivery of DNA to an individual.

3. The method of claim 1, wherein the secondary structure of the mRNA transcribed from a native heterologous DNA sequence is predicted using a computer and computer program.

4. The method of claim 1, wherein the secondary structure of the mRNA transcribed from a native heterologous DNA sequence is modified by increasing the AT content of the coding sequence.

5. The method of claim 4, wherein the secondary structure of the mRNA transcribed from a native heterologous DNA sequence is modified by increasing the AT content of the coding sequence at the 5' end of the coding sequence such that mRNA transcribed therefrom has an increased AU content.

6. The method of claim 5, wherein the secondary structure of the mRNA transcribed from a native heterologous DNA sequence is modified by increasing the AT content of the coding

sequence at the 5' end of the coding sequence within 200 nucleotides from the initiation codon such that mRNA transcribed therefrom has an increased AU content.

7. The method of claim 6 wherein the secondary structure of the mRNA transcribed from a native heterologous DNA sequence is modified by increasing the AT content of the coding sequence at the 5' end of the coding sequence within 150 nucleotides from the initiation codon such that mRNA transcribed therefrom has an increased AU content.

8. The method of claim 6 wherein the secondary structure of the mRNA transcribed from a native heterologous DNA sequence is modified by increasing the AT content of the coding sequence at the 5' end of the coding sequence within 100 nucleotides from the initiation codon such that mRNA transcribed therefrom has an increased AU content.

9. An injectable pharmaceutical composition comprising a nucleic acid molecule that includes a modified coding sequence encoding a protein operably linked to regulatory elements, wherein the modified coding sequence comprises a higher AT or AU content relative to the AT or AU content of the native coding sequence, and further comprising a pharmaceutically acceptable carrier.

10. The injectable pharmaceutical composition of claim 9, wherein said modified coding sequence comprises a higher AT or AU content in the first 200 bases relative to the AT or AU content of the native nucleic acid sequence.

11. The injectable pharmaceutical composition of claim 9, wherein said modified coding sequence comprises a higher AT or AU content in the first 150 bases relative to the AT or AU content of the native nucleic acid sequence.

12. The injectable pharmaceutical composition of claim 9, wherein said modified coding sequence comprises a higher AT or AU content in the first 100 bases relative to the AT or AU content of the native nucleic acid sequence.

13. The injectable pharmaceutical composition of claim 9, wherein said modified coding

sequence comprises a higher AT or AU content in at least one region of up to 200 bases in length relative to the AT or AU content of the native nucleic acid sequence.

14. The injectable pharmaceutical composition of claim 9, wherein said modified coding sequence comprises a higher AT or AU content in at least one region of up to 150 bases in length relative to the AT or AU content of the native nucleic acid sequence.

15. The injectable pharmaceutical composition of claim 9, wherein said modified coding sequence comprises a higher AT or AU content in at least one region of up to 100 bases in length relative to the AT or AU content of the native nucleic acid sequence.

16. The injectable pharmaceutical composition of claim 9, wherein the modified coding sequence encodes an immunogen.

17. The injectable pharmaceutical composition of claim 16, wherein the immunogen is a pathogen derived proteins or immunogenic fragment thereof.

18. The injectable pharmaceutical composition of claim 16, wherein the immunogen is a fusion protein that includes a pathogen derived protein or immunogenic fragment thereof.

19. The injectable pharmaceutical composition of claim 16, wherein the immunogen is a cancer antigen or immunogenic fragment thereof.

20. The injectable pharmaceutical composition of claim 16, wherein the immunogen is a fusion protein that includes a cancer antigen or immunogenic fragment thereof.

21. The injectable pharmaceutical composition of claim 16, wherein the immunogen is an autoimmune disease associated protein or immunogenic fragment thereof.

22. The injectable pharmaceutical composition of claim 16, wherein the immunogen is a fusion protein that includes an autoimmune disease associated protein or immunogenic fragment thereof.

23. The injectable pharmaceutical composition of claim 9, wherein the modified coding sequence encodes a non-immunogenic therapeutic protein.

24. The injectable pharmaceutical composition of claim 23, wherein the non-immunogenic therapeutic protein is selected from the group consisting of cytokines, growth factors, blood products, and enzymes.

25. The injectable pharmaceutical composition of claim 9, wherein the modified coding sequence comprises dispersed modifications.

26. The injectable pharmaceutical composition of claim 25, wherein the dispersed modifications are at least two modified coding sequences of up to 200 bases in length alternating with regions of native coding sequence.

27. The injectable pharmaceutical composition of claim 25, wherein the dispersed modifications are at least two modified coding sequences of up to 150 bases in length alternating with regions of native coding sequence.

28. The injectable pharmaceutical composition of claim 25, wherein the dispersed modifications are at least two modified coding sequences of up to 100 bases in length alternating with regions of native coding sequence.

29. The injectable pharmaceutical composition of claim 9, wherein said modified coding sequence comprises a higher AT or AU content in the last 200 bases relative to the AT or AU content of the native nucleic acid sequence.

30. The injectable pharmaceutical composition of claim 9, wherein said modified coding sequence comprises a higher AT or AU content in the last 150 bases relative to the AT or AU content of the native nucleic acid sequence.

31. The injectable pharmaceutical composition of claim 9, wherein said modified coding sequence comprises a higher AT or AU content in the last 100 bases relative to the AT or AU

content of the native nucleic acid sequence.

32. A recombinant viral vector comprising a nucleic acid molecule that includes a modified coding sequence encoding a protein operably linked to regulatory elements, wherein the modified coding sequence comprises a higher AT or AU content relative to the AT or AU content of the native coding sequence.

33. The recombinant viral vector of claim 32, wherein said modified coding sequence comprises a higher AT or AU content in the first 200 bases relative to the AT or AU content of the native nucleic acid sequence.

34. The recombinant viral vector of claim 32, wherein said modified coding sequence comprises a higher AT or AU content in the first 150 bases relative to the AT or AU content of the native nucleic acid sequence.

35. The recombinant viral vector of claim 32, wherein said modified coding sequence comprises a higher AT or AU content in the first 100 bases relative to the AT or AU content of the native nucleic acid sequence.

36. The recombinant viral vector of claim 32, wherein said modified coding sequence comprises a higher AT or AU content in at least one region of up to 200 bases in length relative to the AT or AU content of the native nucleic acid sequence.

37. The recombinant viral vector of claim 32, wherein said modified coding sequence comprises a higher AT or AU content in at least one region of up to 150 bases in length relative to the AT or AU content of the native nucleic acid sequence.

38. The recombinant viral vector of claim 32, wherein said modified coding sequence comprises a higher AT or AU content in at least one region of up to 100 bases in length relative to the AT or AU content of the native nucleic acid sequence.

39. The recombinant viral vector of claim 32, wherein the modified coding sequence encodes

an immunogen.

40. The recombinant viral vector of claim 39, wherein the immunogen is a pathogen derived proteins or immunogenic fragment thereof.

41. The recombinant viral vector of claim 39, wherein the immunogen is a fusion protein that includes a pathogen derived protein or immunogenic fragment thereof.

42. The recombinant viral vector of claim 39, wherein the immunogen is a cancer antigen or immunogenic fragment thereof.

43. The recombinant viral vector of claim 32, wherein the immunogen is a fusion protein that includes a cancer antigen or immunogenic fragment thereof.

44. The recombinant viral vector of claim 39, wherein the immunogen is an autoimmune disease associated protein or immunogenic fragment thereof.

45. The recombinant viral vector of claim 39, wherein the immunogen is a fusion protein that includes an autoimmune disease associated protein or immunogenic fragment thereof.

46. The recombinant viral vector of claim 32, wherein the modified coding sequence encodes a non-immunogenic therapeutic protein.

47. The recombinant viral vector of claim 46, wherein the non-immunogenic therapeutic protein is selected from the group consisting of cytokines, growth factors, blood products, and enzymes.

48. The recombinant viral vector of claim 32, wherein the modified coding sequence comprises dispersed modifications.

49. The recombinant viral vector of claim 48, wherein the dispersed modifications are at least two modified coding sequences of 200 bases in length alternating with regions of native coding

sequence.

50. The recombinant viral vector of claim 48, wherein the dispersed modifications are at least two modified coding sequences of 150 bases in length alternating with regions of native coding sequence.

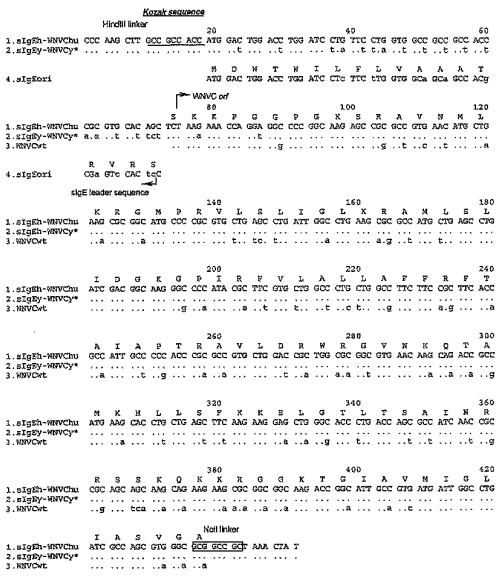
51. The recombinant viral vector of claim 48, wherein the dispersed modifications are at least two modified coding sequences of 100 bases in length alternating with regions of native coding sequence.

52. The recombinant viral vector of claim 32, wherein said modified coding sequence comprises a higher AT or AU content in the last 200 bases relative to the AT or AU content of the native nucleic acid sequence.

53. The recombinant viral vector of claim 32, wherein said modified coding sequence comprises a higher AT or AU content in the last 150 bases relative to the AT or AU content of the native nucleic acid sequence.

54. The recombinant viral vector of claim 32, wherein said modified coding sequence comprises a higher AT or AU content in the last 100 bases relative to the AT or AU content of the native nucleic acid sequence.

Fig.1.


Kozak sequence
HindIII linker
WNVC orf
sigE leader sequence
NotI linker

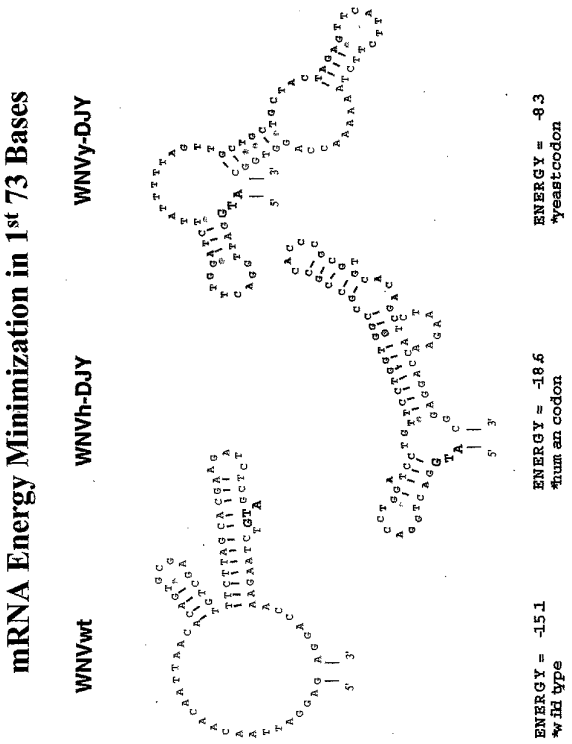

mRNA Energy Minimization in 1st 73 Bases
WNVwt
WNVh-DJY
WNVy-DJY
ENERGY = -15.1
*wild type
ENERGY = -18.6
*human codon
ENERGY = -8.3
*yeastcodon

Fig. 2

Fig. 3 Immunoprecipitation of Radiolabeled *In Vitro* Translated Proteins

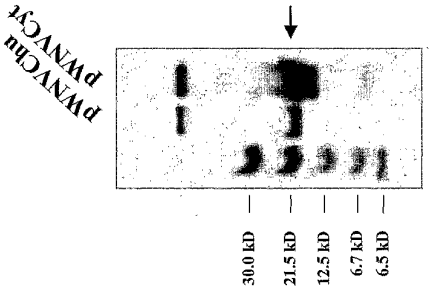
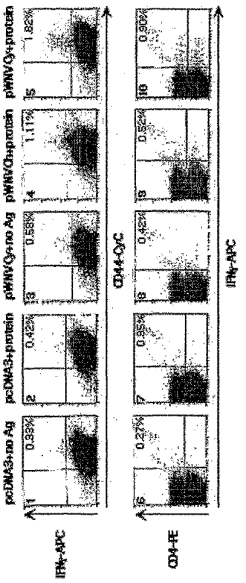
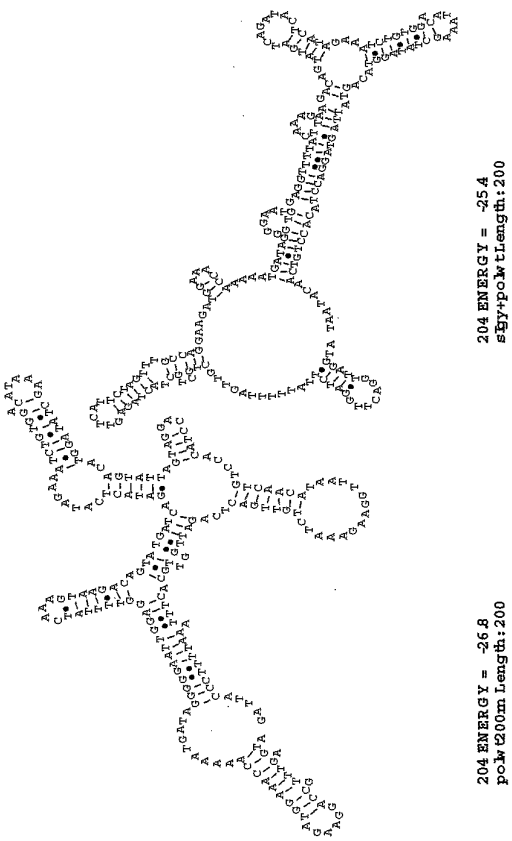


Fig. 4


pcDNA3-no Ag
pcDNA3-protein
pWNVCy-no Ag
pWNVCh-protein
pWNVCy-protein
IFN-APC
CD4-PE
CD44-CyC
0.39%
0.42%
0.55%
1.17%
1.82%
0.27%
0.85%
0.42%
0.52%
0.80%

mRNA Energy Minimization: Addition of Leader Sequence

Fig. 5



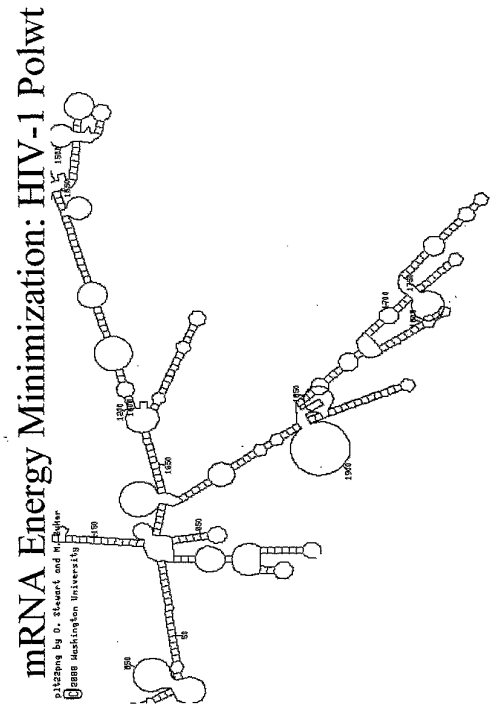

mRNA Energy Minimization: HIV-1 Polwt
p1x2png by D. Stewart and M. Zuker
©2000 Washington University

Fig. 6

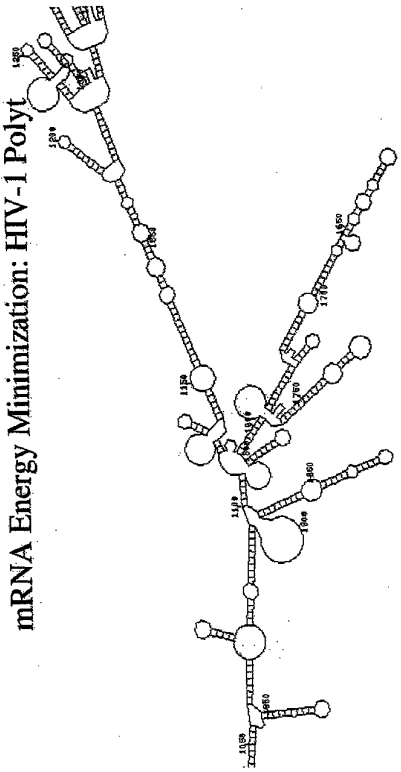
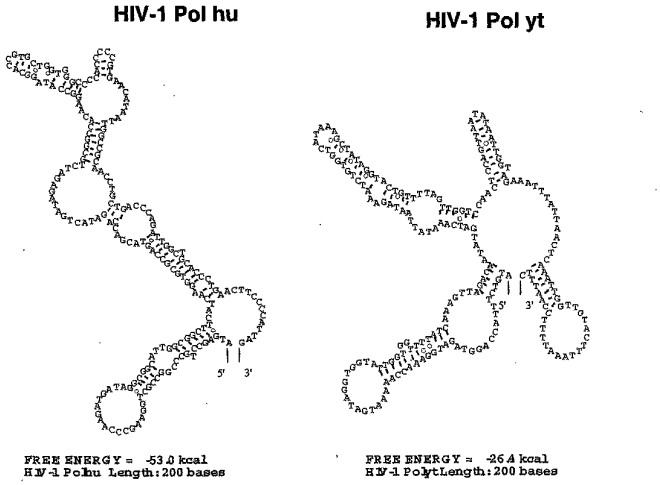


Fig 1

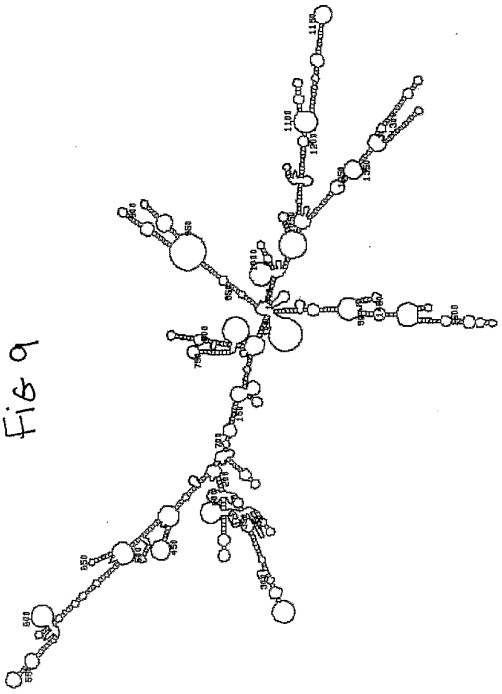
FIG 8

mRNA Energy Minimization with 1st 200 bases



mRNA Energy Minimization: HIV-1 Gag-wt

Fig 9



mRNA Energy Minimization: HIV-1 Gag-yt

Fig 10

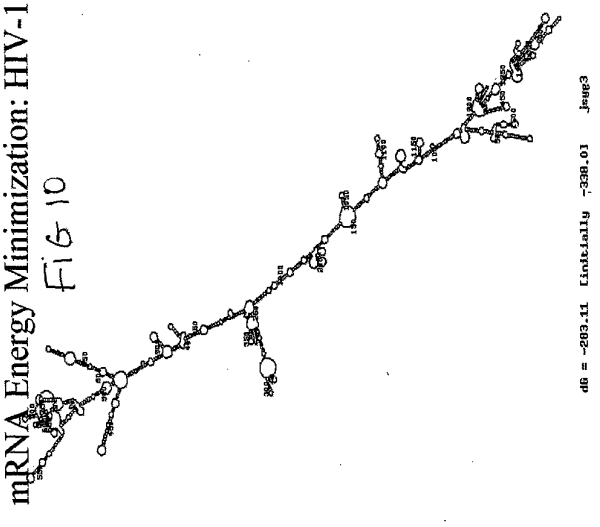
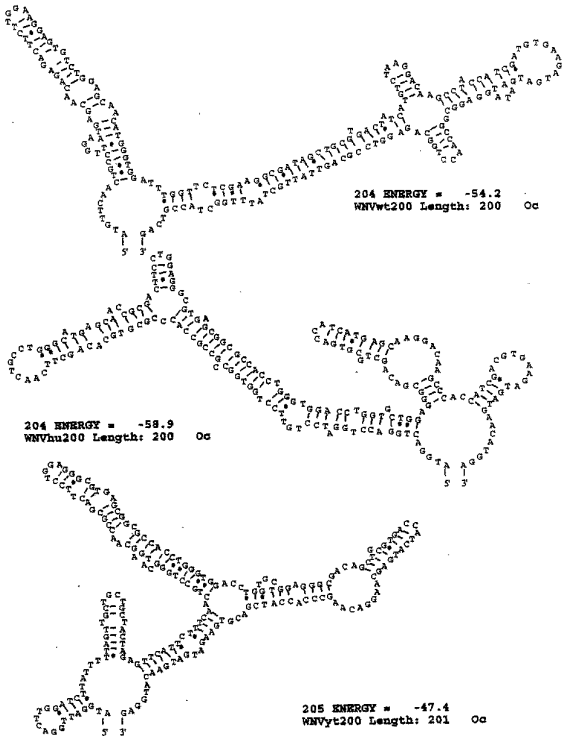


Fig 11

WNV Env RNA Optimization



【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)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
11 April 2002 (11.04.2002)

PCT

(10) International Publication Number
WO 02/029088 A3

(51) International Patent Classification⁷: C12P 21/00, 21/02, A01N 43/04, C12N 15/00, 15/09, 15/11, 15/63

(21) International Application Number: PCT/US01/31451

(22) International Filing Date: 4 October 2001 (04.10.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/237,885 4 October 2000 (04.10.2000) US

(71) Applicant *(for all designated States except US)*: THE TRUSTEES OF THE UNIVERSITY OF PENNSYLVANIA [US/US]; Suite 300, 3700 Market Street, Philadelphia, PA 19104-3147 (US).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: WEINER, David, B. [US/US]; 717 Beacon Lane, Merion, PA 19066 (US). YANG, Joo-Sung [KR/US]; 102 South 42nd Street, Apartment 1, Philadelphia, PA 19104 (US).

(74) Agents: DeLUCA, Mark et al.; Woodcock Washburn LLP, 46th floor, One Liberty Place, Philadelphia, PA 19103 (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, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, 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:
— *with international search report*

(88) Date of publication of the international search report:
18 July 2002

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 02/029088 A3

(54) Title: HIGHLY EXPRESSIBLE GENES

(57) Abstract: The present invention provides methods of producing protein in a recombinant expression system that comprises translation of mRNA transcribed from a heterologous DNA sequence in the expression system, said method comprising the steps of predicting the secondary structure of mRNA transcribed from a native heterologous DNA sequence wherein mRNA transcribed from the modified heterologous DNA sequence has a secondary structure having increased free energy compared to that of the secondary structure of the mRNA transcribed from the native heterologous DNA sequence; and using the modified heterologous DNA sequence in the recombinant expression system for protein production. The invention also provides injectable pharmaceutical compositions comprising a nucleic acid molecule that includes a modified coding sequence. The invention also provides coding sequence.

【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
11 April 2002 (11.04.2002)

PCT

(10) International Publication Number
WO 02/029088 A3

(51) International Patent Classification⁷: C12P 21/00, 21/02, A01N 43/04, C12N 15/00, 15/09, 15/11, 15/63

(21) International Application Number: PCT/US01/31451

(22) International Filing Date: 4 October 2001 (04.10.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/237,885 4 October 2000 (04.10.2000) US

(71) Applicant *(for all designated States except US)*: **THE TRUSTEES OF THE UNIVERSITY OF PENNSYLVANIA** [US/US]; Suite 300, 3700 Market Street, Philadelphia, PA 19104-3147 (US).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: **WEINER, David, B.** [US/US]; 717 Beacon Lane, Merion, PA 19066 (US). **YANG, Joo-Sung** [KR/US]; 102 South 42nd Street, Apartment 1, Philadelphia, PA 19104 (US).

(74) Agents: **DeLUCA, Mark** et al.; Woodcock Washburn LLP, 46th floor, One Liberty Place, Philadelphia, PA 19103 (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, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, 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:
— *with international search report*

(88) Date of publication of the international search report:
18 July 2002

(48) Date of publication of this corrected version:
20 February 2003

(15) Information about Correction:
see PCT Gazette No. 08/2003 of 20 February 2003, 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 02/029088 A3

(54) Title: HIGHLY EXPRESSIBLE GENES

(57) Abstract: The present invention provides methods of producing protein in a recombinant expression system that comprises translation of mRNA transcribed from a heterologous DNA sequence in the expression system, said method comprising the steps of predicting the secondary structure of mRNA transcribed from a native heterologous DNA sequence wherein mRNA transcribed from the modified heterologous DNA sequence has a secondary structure having increased free energy compared to that of the secondary structure of the mRNA transcribed from the native heterologous DNA sequence; and using the modified heterologous DNA sequence in the recombinant expression system for protein production. The invention also provides injectable pharmaceutical compositions comprising a nucleic acid molecule that includes a modified coding sequence. The invention also provides coding sequence.

HIGHLY EXPRESSIBLE GENES

CROSS-REFERENCE TO RELATED APPLICATIONS

This application claims benefit under 35 U.S.C. §119(e) of U.S. Provisional Patent Application Serial No. 60/237,885, filed October 4, 2000, incorporated herein by reference.

FIELD OF THE INVENTION

The present invention relates generally to the fields of gene expression, gene therapy, and genetic immunization.

BACKGROUND OF THE INVENTION

The expression of a protein gene product is influenced by many factors, including gene copy number, gene integration site or gene location in the genome, transcription factors, mRNA stability, and translation efficiency. For example, the expression of the human immunodeficiency virus-1 (HIV-1) structural genes *gag*, *pol*, and *env* is dependent on the Rev/Rev-responsive element (RRE) at a posttranscriptional level. This dependency on Rev is a limiting factor for gene expression. In addition, highly stable RNA secondary structures that form in various regions of the HIV RNA transcript can block or otherwise interfere with ribosome movement, and thus effectively limit translation. Formation of stable RNA secondary structures in gene transcripts is a general phenomenon that can limit the translational yield of many protein gene products for a wide variety of genes.

Kim *et al.*, 1997, Gene, 199:293-301, which is incorporated herein by reference, optimized expression of human erythropoietin (EPO) in mammalian cells by altering the codons encoding the leader sequence and the first 6 amino acids of the mature EPO protein for the most

prevalently used yeast codons, and changing the codons encoding the rest of the EPO protein for the most prevalently used human codons.

U.S. patents US 5,972,596 and 5,965,726 (Pavlakis *et al.*), which are incorporated herein by reference, describe methods of locating an inhibitory/instability sequence or sequences (INS: sequences that render an mRNA unstable or poorly utilized/translated) within the coding region of an mRNA and modifying the gene encoding the mRNA to remove the inhibitory/instability sequences with clustered nucleotide substitutions.

There is a need for new methods of expressing proteins and methods of increasing the level of protein expression of therapeutic and immunogenic transgenes. There is a need for methods of increasing the translational yields of any protein gene product. There is a need for methods of overcoming the limitations imposed by RNA secondary structure in RNA transcripts upon the ultimate level of protein expression of any gene. The present invention is directed to addressing these and other needs.

SUMMARY OF THE INVENTION

The present invention provides methods of producing protein in a recombinant expression system that comprises translation of mRNA transcribed from a heterologous DNA sequence in the expression system, said method comprising the steps of predicting the secondary structure of mRNA transcribed from a native heterologous DNA sequence; modifying the native heterologous DNA sequence to produce a modified heterologous DNA sequence wherein mRNA transcribed from the modified heterologous DNA sequence has a secondary structure having increased free energy compared to that of the secondary structure of the mRNA transcribed from the native heterologous DNA sequence; and using the modified heterologous DNA sequence in the recombinant expression system for protein production. The recombinant expression system may be a cell free *in vitro* transcription and translation system, an *in vitro* cell expression system, a DNA construct used in direct DNA injection, or a recombinant vector for delivery of DNA to an individual. The secondary structure of the mRNA transcribed from a native heterologous DNA sequence may be predicted using a computer and computer program. The native heterologous DNA sequence may be modified by increasing the AT content of the coding sequence, in particular, at the5' end of the coding sequence, or at the 5' end of the coding sequence within 200, 150, or 100 nucleotides from the initiation codon.

The present invention also provides injectable pharmaceutical compositions comprising a nucleic acid molecule that includes a modified coding sequence encoding a protein operably linked to regulatory elements, wherein the modified coding sequence comprises a higher AT or AU content relative to the AT or AU content of the native coding sequence, and further comprising a pharmaceutically acceptable carrier. The encoded proteins may be immunogens or non-immunogenic therapeutic proteins. The modifications may be within the first 100 to 200 bases of the coding sequence, within stretches of sequences dispersed throughout the coding sequence, or within in the last 100 to 200 bases.

The present invention also provides recombinant viral vectors comprising a nucleic acid molecule that includes a modified coding sequence encoding a protein operably linked to regulatory elements, wherein the modified coding sequence comprises a higher AT or AU content relative to the AT or AU content of the native coding sequence.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 presents the nucleotide and amino acid sequence of the West Nile Virus (WNV) wild type capsid (Cp) protein (WNVC) with constructs (WNVChu and WNVCy*) modified on the basis of RNA secondary structure. A secretory IgE signal leader sequence was fused to the WNVC protein. To avoid varied expression levels due to the linear sequence between the promoter and 5'-proximal region of the WNVC, the leader sequences and the codons for amino acids 2 - 6 of the WNVC were modified with yeast (WNVCy) or human (WNVChu) optimized codons. However, the remaining portion of the coding sequence for the WNV capsid protein, in both constructs, was modified with human optimized codons. Presented are 1) the wild type nucleotide sequence encoding the sIgE leader sequence (4. sIgEori) (SEQ ID NO:1), 2) the amino acid sequence of the sIgE leader sequence (appearing above the nucleotide sequence) (SEQ ID NO:2), 3) the amino acid sequence for the WNV capsid protein (minus the initial methionine) (SEQ ID NO:3), and 4) the nucleotide sequence of the sIgEh-WNV capsid protein encoding sequence of the WNVChu construct (1. sIgEh-WNVChu) (SEQ ID NO:4). Differences in the coding sequence for sIgEh-WNV capsid protein in the WNVCy construct (2. sIgEy-WNVCy*) and in the wild type WNV capsid encoding sequence (3. WNVCwt) are indicated below the nucleotide sequence of the WNVChu construct.

Figure 2 presents the MulFold predicted RNA secondary structures with free energy values for the first 73 nucleotides of 1) the wild type mRNA encoding WVN capsid protein

(WNVwt), 2) an mRNA encoding the sIgE leader/WNV capsid protein containing human optimized codons (WNVh-DJY), and 3) an mRNA encoding the sIgE leader/WNV capsid protein containing yeast optimized codons (WNVy-DJY). The last codon (GGC for glycine) shown for the WNVy-DJY sequence is human optimized. As shown, "T" represents "U" in the RNA strands. The nucleotides of the mRNA strands that encode the sIgE leader portion of the fusions in WNVh-DJY and WNVy-DJY are shown in bold.

Figure 3 presents an autoradiograph of electrophoretically separated, immunoprecipitated, radiolabeled *in vitro* transcription/translation products from two different WNV capsid protein constructs: pWNVChu (also called WNVChu and pWNVh-DJY) and pWNVCyt (also called WNVCy and pWNVy-DJY). The first lane on the left contains molecular weight markers. The arrow indicates the position of the main capsid protein product. The proteins, which are fusions with polyhistidine C-terminal tags, were immunoprecipitated using an anti-His antibody.

Figure 4 presents the flow cytometry analysis of intracellular IFN-γ expression in *in vitro* stimulated splenocytes from DNA immunized mice. Values presented are the percentage dual positive cells. In the upper panels, the cells were stained for INF-γ and CD44; in the lower panels the cells were stained for CD4 and IFN-γ. The labeling across the top indicates the vector used to immunize the mice plus the stimulus used for the *in vitro* restimulation of the splenocytes. The immunizing vectors were pcDNA3 (pcDNA3.1), pWNVh-DJY (pWNVCh), and pWNVy-DJY (pWNVCy). "No Ag" indicates that the splenocytes were incubated with an *in vitro* translation control (described in Example 2), "protein" indicates that the splenocytes were incubated with *in vitro* translated Cp protein product from the pWNVy-DJY expression construct.

Figure 5 presents the MulFold predicted RNA secondary structure with free energy values based upon energy minimization for the first 200 nucleotides of the wild type mRNA for to the HIV-1 *pol* gene (polwt200rn) and for the fist 200 nucleotides of an mRNA for HIV-1 *pol* gene including a 5' sequence encoding the IgE leader sequence with codons less prevalently used in humans (yeast optimized) (sIgy+polwt). As shown, "T" represents "U" in the RNA strand.

Figure 6 presents the MulFold predicted secondary structure of the mRNA for the HIV-1 *pol* structural gene.

Figure 7 presents the MulFold predicted secondary structure for the mRNA for the HIV-1 *pol* structural gene after the 200 nucleotide region of the sequence from nucleotide 1738 through

nucleotide 1938 has been altered to contain codons that are less prevalently utilized in humans (yeast optimized codons).

Figure 8 presents the MulFold predicted secondary structure and overall free energy value for the first 200 nucleotides of the mRNA for the HIV-1 *pol* gene containing human optimized codons (HIV-1 Pol hu), and for the mRNA for the HIV-1 *pol* gene containing codons less prevalently utilized in humans (yeast optimized codons) (HIV-1 Pol yt). As shown, "T" represents "U" in the RNA strands.

Figure 9 presents the MulFold predicted secondary structure and overall free energy value for the mRNA transcript for the HIV-1 *gag* structural gene.

Figure 10 presents the MulFold predicted secondary structure and overall free energy value for the mRNA transcript for the HIV-1 *gag* structural gene altered with codons that are utilized less prevalently in humans (yeast optimized).

Figure 11 presents the MulFold predicted secondary structures and overall free energy values for the first 200 nucleotides of the mRNA transcript for 1) the wild type West Nile Virus (WNV) envelope (*env*) gene (WNVwt200), 2) the WNV *env* gene optimized with the most prevalently used codons in humans (WNVhu200), and 3) the WNV *env* gene having codons that are utilized less prevalently in humans (yeast optimized, WNVyt200). As shown, "T" represents "U" in the RNA strands.

DESCRIPTION OF THE PREFERRED EMBODIMENTS

The present invention is based upon the discovery that enhancement of protein expression can be achieved by increasing the free energy of and destabilizing RNA secondary structure through changes at the nucleotide level. It has been discovered that an increase in the free energy (X kcal) of an RNA transcript will result in increased expression of the protein that it encodes. In preferred embodiments, an increase in the free energy (X kcal) within a 200 base segment of an RNA transcript will result in increased expression of the protein that it encodes. The segment is preferably at the 5' end, usually including the initiation codon. In some embodiments, the segment is preferably 200 bases, 150 bases, or 100 bases. The secondary structure of an RNA molecule is the collection of base pairs that occur in its three-dimensional structure. The secondary structure of a given RNA molecule can be predicted and such predicted secondary structure will have an assigned overall free energy value. It has been discovered that alterations to the primary sequence of an RNA transcript that result in an increase to the minimum predicted

overall free energy for a predicted secondary structure for that RNA, or that increase the minimum predicted free energy for a predicted secondary structure for regions of that RNA, will promote increased expression of the protein encoded by that RNA transcript. This strategy for the optimization of protein expression applies to any situation where expression is desired, including, but not limited to: *in vivo*, including, but not limited to, DNA vaccines, live vaccines, gene therapeutics, and transgenes; *in vitro*, including, but not limited to, recombinant manufacturing procedures using such systems as prokaryotic and eukaryotic (mammal, insect, and yeast) cells in culture; *ex vivo*, including, but not limited to, systems where cells receive expression constructs and are implanted into recipient organisms; and any other expression system where it is desirable to express a gene of interest or increase the expression of a gene.

One aspect of the invention is to generate an RNA encoding a protein that promotes efficient expression of that protein or that leads to increased levels of expression of the protein. Alterations to the sequence of the DNA encoding the RNA that lead to an increase in the minimum overall free energy for the predicted secondary structure of that RNA, or that increase the minimum free energy for the predicted secondary structure of one or more regions of that RNA promote efficient and/or increased expression of the encoded protein.

Increases to the free energy of the secondary structure of an RNA can be monitored by analyzing various altered versions of a sequence with a program like MFOLD, which calculates and predicts the most stable structure for an input sequence based upon energy minimization. MFOLD is computer software designed by Zuker, Jaeger, and colleagues (*see* Zuker, 1989, On finding all suboptimal foldings of an RNA molecule, Science, 244:48-52, and Jaeger *et al.*, 1989, Improved predictions of secondary structures for RNA, Proc. Natl. Acad. Sci. USA, 86:7706-7710, each of which is incorporated herein by reference) that is used for the prediction of RNA secondary structure by free energy minimization, using energy rules developed by Turner and colleagues (*see* Freier *et al.*, 1986, Proc. Natl. Acad. Sci. USA, 83:9373-9377, which is incorporated herein by reference). MulFold is the Macintosh version of MFOLD. LoopDloop is a secondary structure drawing program. The most stable structure will be the one with a minimum overall free energy. The more negative the value of the free energy for the structure, the more stable. Alterations to the sequence of the RNA that are predicted to result in a secondary structure having an overall higher free energy value, are destabilizing alterations which result in less stable RNA secondary structure and which promote efficient translation of the RNA and an increase in protein expression.

The practice of the present invention will employ, unless otherwise indicated, conventional methods of virology, immunology, microbiology, molecular biology, and recombinant DNA techniques within the skill of the art. Such techniques are explained fully in the literature. See, *e.g.*, Sambrook *et al.*, eds., Molecular Cloning: A Laboratory Manual (3rd ed.) Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY (2001); Ausubel *et al.*, eds., Current Protocols in Molecular Biology, John Wiley & Sons, New York, NY (2001); Glover & Hames, eds., DNA Cloning 3: A Practical Approach, Vols. I, II, & III, IRL Press, Oxford (1995); Colowick & Kaplan, eds., Methods in Enzymology, Academic Press; Weir *et al.*, eds., Handbook of Experimental Immunology, 5th ed., Blackwell Scientific Publications, Ltd., Edinburgh, (1997); Fields, Knipe, & Howley, eds., Fields Virology (3rd ed.) Vols. I & II, Lippincott Williams & Wilkins Pubs. (1996); Flint, *et al.*, eds., Principles of Virology: Molecular Biology, Pathogenesis, and Control, ASM Press, (1999); Coligan *et al.*, eds., Current Protocols in Immunology, John Wiley & Sons, New York, NY (2001), each of which is incorporated herein by reference.

Various definitions are made throughout this document. Most words have the meaning that would be attributed to those words by one skilled in the art. Words specifically defined either below or elsewhere in this document have the meaning provided in the context of the present invention as a whole and as typically understood by those skilled in the art.

As used herein, the term "recombinant expression system" refers to any nucleic acid based approach or system for the expression of a gene product or gene products of interest, that has been artificially organized (man made) of components directed toward the expression of the gene product or products. The components may be of naturally occurring genetic sources, synthetic or artificial, or some combination of natural and artificial genetic elements. Generally the gene product is a protein, polypeptide, or peptide. Examples of recombinant expression systems include, but are not limited to, a cell free *in vitro* transcription and translation system; an *in vitro* cell expression system; a DNA construct used in direct DNA injection; and a recombinant vector for delivery of DNA to an individual.

As used herein, the term "heterologous DNA sequences" refers to deoxyribonucleic acid based sequences that are in a non-natural context, for example, in a recombinant construct, plasmid, or virus, or inserted into a non-natural position in a chromosome, or introduced into a non-natural or foreign cell. "Heterologous DNA sequence" refers to any DNA sequence that is foreign or not naturally associated with the other DNA sequences to which it is associated or

linked (operably or otherwise), or a DNA sequence that is not naturally associated with the cell or organism into which it is introduced. An example of a heterologous DNA sequence is one that is used for the expression of a foreign or heterologous protein gene product in a host cell or organism. A heterologous DNA sequence can also be a part of a vector or expression construct having genetic material designed for directing the expression of a gene product, such as a protein, polypeptide, or peptide, in a host cell *in vivo* or *in vitro*, or in a cell free *in vitro* expression system.

As used herein, the term "native heterologous DNA sequence" refers to a heterologous DNA sequence that, although positioned in a non-natural context, has a nucleotide sequence that is not modified or altered from the sequence it has in its natural context. For example, a viral gene may be inserted into a recombinant expression construct, such that the viral gene is a heterologous DNA sequence with respect to other sequences in the construct, but without introduction of any changes to the nucleotide sequence of the viral gene. In this example, as a native heterologous DNA sequence, the viral gene has the native nucleotide sequence as would be found in its natural context within the genome of the virus, but the viral gene sequence is heterologous with respect to its new context. A native heterologous DNA sequence can also be any DNA sequence that is considered to be the reference or starting version of a DNA sequence, from which a modified (non-native) version of the DNA sequence, containing alterations to the nucleic acid sequence may be prepared. A native heterologous DNA sequence can also be composed of multiple native DNA sequences that are unaltered in sequence from that which is found in nature, but that are not naturally found together. An example of such a native heterologous DNA sequence composed of multiple native DNA sequences is a fusion gene composed of native genetic sequence from two different genes.

As used herein, the term "modified heterologous DNA sequence" refers to a heterologous DNA sequence that is not only positioned in a non-natural context, but also has a nucleotide sequence that is modified or altered from the sequence it has in its natural context. For example, a viral gene that is a modified heterologous DNA sequence will be inserted into a recombinant expression construct, such that the viral gene is heterologous with respect to other sequences in the construct, and further, will have a nucleotide sequence that is modified or altered and not the native nucleotide sequence as found in its natural context within the genome of the virus.

As used herein, the term "increased free energy" in reference to RNA secondary structure, refers to an increase in the free energy value for an RNA secondary structure. Free energy values that are more negative are lower than values that are less negative.

As used herein, the term "modified coding sequence" refers to a nucleic acid sequence (DNA- or RNA-based), that encodes a gene product, protein, polypeptide, or peptide, and that has been modified or altered from the native or naturally-occurring coding sequence for that gene product, protein, polypeptide, or peptide. The coding sequence may be comprised of sequences from more than one genetic source, for example, the coding sequence may be a fusion gene encoding a fusion protein having a leader sequence from a gene for one protein and the remaining sequence from a gene for another protein, brought together as one hybrid coding sequence, that is non-natural. In the case of such an example of a coding sequence comprised of sequences from more than one genetic source, "modified coding sequence" indicates that any modification is relative to the native or naturally-occurring coding sequence for the respective separate sequences.

As used herein, the term "native coding sequence" refers to a nucleic acid sequence (DNA- or RNA-based), that encodes a gene product, protein, polypeptide, or peptide, and that has not been modified or altered from the native or naturally-occurring coding sequence for that gene product, protein, polypeptide, or peptide. If the coding sequence encodes a fusion protein, the component parts have not been modified or altered from the native or naturally-occurring coding sequence for those component parts.

As used herein, the term "higher AT or AU content" refers to modifications to a coding sequence which render it a "modified coding sequence" such that if it is DNA-based it has a higher concentration of adenine and thymidine residues than the corresponding native coding sequence, and if it is RNA-based it has a higher concentration of adenine and uridine residues than the corresponding native coding sequence.

As used herein, the term "the first 200 bases" in reference to a modified coding sequence that has been modified relative to the native coding sequence, refers to the first 200 contiguous nucleotide bases from the 5' end of the respective coding sequence.

As used herein, the term "the first 150 bases" in reference to a modified coding sequence that has been modified relative to the native coding sequence, refers to the first 150 contiguous nucleotide bases from the 5' end of the respective coding sequence.

As used herein, the term "the first 100 bases" in reference to a modified coding sequence that has been modified relative to the native coding sequence, refers to the first 100 contiguous nucleotide bases from the 5' end of the respective coding sequence.

As used herein, the term "the last 200 bases" in reference to a modified coding sequence that has been modified relative to the native coding sequence, refers to the last 200 contiguous nucleotide bases from the 3' end of the respective coding sequence.

As used herein, the term "the last 150 bases" in reference to a modified coding sequence that has been modified relative to the native coding sequence, refers to the last 150 contiguous nucleotide bases from the 3' end of the respective coding sequence.

As used herein, the term "the last 100 bases" in reference to a modified coding sequence that has been modified relative to the native coding sequence, refers to the last 100 contiguous nucleotide bases from the 3' end of the respective coding sequence.

As used herein, the term "region of up to 200 bases in length" in reference to a coding sequence, refers to a region of up to 200 contiguous nucleotide bases of the coding sequence. The region may be anywhere within the coding sequence.

As used herein, the term "region of up to 150 bases in length" in reference to a coding sequence, refers to a region of up to 150 contiguous nucleotide bases of the coding sequence. The region may be anywhere within the coding sequence.

As used herein, the term "region of up to 100 bases in length" in reference to a coding sequence, refers to a region of up to 100 contiguous nucleotide bases of the coding sequence. The region may be anywhere within the coding sequence.

As used herein, the term "dispersed modifications" refers to any combination of at least two regions of contiguous nucleotide bases that are modified to have a higher AT or AU content relative to the native coding sequence in the respective regions, and that are dispersed throughout the sequence such that regions of modified coding sequences will alternate with regions of native coding sequence. By way of non-limiting example, a modified coding sequence may contain alternating regions of modifications, wherein the first 200 contiguous bases of the coding sequence have a higher AT or AU content relative to the native coding sequence, the next 200 bases of the coding sequence are non-modified relative to the native coding sequence, and the subsequent 200 contiguous base region is modified to have a higher AT or AU content relative to the native coding sequence. The size of the modified regions may be of any length, and is preferably 200, 150, or 100 bases in length. The size of non-modified regions will be of variable

length depending on the positioning of the modified regions. In preferred embodiments the dispersed modifications comprise alternating regions of modified and native coding sequence over the entire coding sequence, where the size of each alternating region is preferably 200 or 150 or 100 bases in length.

As used herein, "injectable pharmaceutical composition" refers to pharmaceutically acceptable compositions for use in patients that are sterile, pyrogen-free, and essentially free of any particulates or particulate matter. *See, Remington's Pharmaceutical Sciences*, 18th Ed., Gennaro, ed., Mack Publishing Co., Easton, PA, 1990 and U.S.P., the standards of the U. S. Pharmacopeia, which is incorporated herein by reference.

As used herein, "pharmaceutically acceptable carrier" includes any carrier that does not itself induce a harmful effect to the individual receiving the composition. For example, a "pharmaceutically acceptable carrier" should not induce the production of antibodies harmful to the recipient. Suitable "pharmaceutically acceptable carriers" are known to those of skill in the art and are described in *Remington's Pharmaceutical Sciences, supra.*

As used herein the term "target protein" is meant to refer to peptides and proteins encoded by gene constructs of the present invention which act as target proteins for an immune response. The terms "target protein" and "immunogen" are used interchangeably and refer to a protein against which an immune response can be elicited. The target protein is an immunogenic protein which shares at least an epitope with a protein from the pathogen or undesirable cell-type such as a cancer cell or a cell involved in autoimmune disease against which an immune response is desired. The immune response directed against the target protein will protect the individual against and/or treat the individual for the specific infection or disease with which the target protein is associated.

As used herein the term "desired protein" is meant to refer to peptides and proteins encoded by gene constructs of the present invention which either act as target proteins for an immune response or as a therapeutic or compensating protein in gene therapy regimens.

As used herein, the phrase "immunogenic fragment thereof" in reference to an immunogen, refers to fragments of less than the full length of the immunogen against which an immune response can be induced.

As used herein, the term "cancer antigens" refers to any proteins, polypeptides, or peptides, and the like, that are associated with and/or serve as markers for cancer, tumors, or cancer cells.

As used herein, the term "autoimmune disease associated proteins" refers to any proteins, polypeptides, or peptides, and the like, that are associated with and/or serve as markers for cells involved in and/or responsible for an autoimmune disease.

As used herein, the term "non-immunogenic therapeutic protein" refers to such proteins, polypeptides, and peptides that are useful for therapeutic treatment of various diseases and disorders, and to which an immune response is not desired and/or not expected upon their introduction into the body of a recipient organism, patient, or individual in need of such therapy or treatment. Examples of "non-immunogenic therapeutic proteins" are proteins that are missing or in low concentration in an individual having a genetic defect in the endogenous gene encoding the protein. Examples of "non-immunogenic therapeutic proteins" include, but are not limited to, cytokines, growth factors, blood products, and enzymes.

As used herein, the term "recombinant viral vector" refers to a construct, based upon the genome of a virus, that can be used as a vehicle for the delivery of nucleic acids encoding proteins, polypeptides, or peptides of interest. Recombinant viral vectors are well known in the art and are widely reported. Recombinant viral vectors include, but are not limited to, retroviral vectors, adenovirus vectors, and adeno-associated virus vectors, which are prepared using routine methods and starting materials.

As used herein, the term "genetic construct" refers to the DNA or RNA molecules that comprise a nucleotide sequence which encodes a target protein or immunomodulating protein. The coding sequence includes initiation and termination signals operably linked to regulatory elements including a promoter and polyadenylation signal capable of directing expression in the cells of the individual to whom the nucleic acid molecule is administered.

As used herein, the term "expressible form" refers to gene constructs which contain the necessary regulatory elements operably linked to a coding sequence that encodes a target protein or an immunomodulating protein, such that when present in the cell of the individual, the coding sequence will be expressed.

As used herein, the term "sharing an epitope" refers to proteins which comprise at least one epitope that is identical to or substantially similar to an epitope of another protein.

As used herein, the term "substantially similar epitope" is meant to refer to an epitope that has a structure which is not identical to an epitope of a protein but nonetheless invokes a cellular or humoral immune response which cross reacts to that protein.

As used herein, the term "intracellular pathogen" is meant to refer to a virus or pathogenic organism that, during at least part of its reproductive or life cycle, exists within a host cell and therein produces or causes to be produced, pathogen proteins.

As used herein, the term "hyperproliferative diseases" is meant to refer to those diseases and disorders characterized by hyperproliferation of cells.

As used herein, the term "hyperproliferative-associated protein" is meant to refer to proteins that are associated with a hyperproliferative disease and/or hyperproliferative cells.

In some preferred embodiments, it is preferred that the alterations to the RNA do not alter the sequence of the protein. In some preferred embodiments, it is preferred that the 200 bases, within which the alterations are introduced, are at the 5' end of the RNA transcript. In some embodiments, it is preferred to increase the free energy in more than one segment of the RNA transcript. Optionally, a leader sequence may be added to increase the free energy of the secondary structure of the RNA.

A stable RNA secondary structure at the 5' end of open reading frame (orf) sequences may block efficient transcription by interfering with ribosome function. Many RNAs have highly stable secondary structural integrity, and these interactions can inhibit gene expression Addition of a sequence encoding a leader, modified such that it was optimized with an AT-rich sequence, resulted in a higher free energy for the predicted RNA structure and allowed efficient initiation by the cellular ribosomes. The stable RNA secondary structure is removed by increasing the free energy.

Therefore, according to the present invention, increasing the AU content in a coding sequence optimizes the sequence by reducing the corresponding RNA secondary structure's integrity, and resulting in increased protein expression/translation, by melting of the inhibitory secondary structures (stem loops) in the RNA transcript. The disruption of secondary structure integrity is particularly important in the 5' portion of the RNA or coding sequence, particularly the first 100 to 200 nucleotides of the RNA. In some embodiments, the AU or AT content is increased in the first 100 to 200 nucletides from the initiation of transcription, and in some embodiments the AU or AT content is increased in the first 100 to 200 nucleotides of the coding sequence or start of translation. In some embodiments, the disruption of secondary structure integrity of the RNA is achieved by full gene changes or alternating patterns within 100 to 200 nucleotide base stretches. Modification of the 3' end is also important.

The strategy of adding a leader-encoding sequence and altering the codons of that sequence to be yeast optimized (less frequently used codons in humans) is applicable to any gene encoding any protein, for example genes encoding viral proteins, including, but not limited to, the HIV-1 *pol* gene. AU-rich content is preferred; human dominant codons/high GC content is not preferred. It has been discovered that lowering the stability of regions of secondary structure within mRNAs can be accomplished without prior knowledge of protein expression or structure. The resultant increased minimum free energy of the secondary structure that is predicted to form from the altered transcript renders the altered transcript capable of enhanced protein expression over the original.

Using standard techniques and readily available starting materials, a modified nucleic acid molecule may be prepared. The nucleic acid molecule may be incorporated into an expression vector which is then incorporated into a host cell. Host cells for use in well known recombinant expression systems for production of proteins are well known and readily available. Examples of host cells include bacteria cells such as *E. coli,* yeast cells such as *S. cerevisiae,* insect cells such as *S. frugiperda,* non-human mammalian tissue culture cells Chinese hamster ovary (CHO) cells and human tissue culture cells such as HeLa cells.

In some embodiments, for example, one having ordinary skill in the art can, using well known techniques, insert DNA molecules into a commercially available expression vector for use in well known expression systems. For example, the commercially available plasmid pSE420 (Invitrogen, San Diego, CA) may be used for production of immunomodulating proteins in *E. coli.* The commercially available plasmid pYES2 (Invitrogen, San Diego, CA) may, for example, be used for production in *S. cerevisiae* strains of yeast. The commercially available MAXBAC™ complete baculovirus expression system (Invitrogen, San Diego, CA) may, for example, be used for production in insect cells. The commercially available plasmid pcDNAI or pcDNA3 (Invitrogen, San Diego, CA) may, for example, be used for production in mammalian cells such as CHO cells. One having ordinary skill in the art can use these commercial expression vectors and systems or others to produce immunomodulating proteins by routine techniques and readily available starting materials. (*See e.g.*, Sambrook *et al.*, eds., 2001, *supra*) Thus, the desired proteins can be prepared in both prokaryotic and eukaryotic systems, resulting in a spectrum of processed forms of the protein.

One having ordinary skill in the art may use other commercially available expression vectors and systems or produce vectors using well known methods and readily available starting

materials. Expression systems containing the requisite control sequences, such as promoters and polyadenylation signals, and preferably enhancers, are readily available and known in the art for a variety of hosts (*See e.g.*, Sambrook *et al.*, eds., 2001, *supra*).

The expression vector including the modified DNA is used to transform the compatible host which is then cultured and maintained under conditions wherein expression of the foreign DNA takes place. The protein of the present invention thus produced is recovered from the culture, either by lysing the cells or from the culture medium as appropriate and known to those in the art. One having ordinary skill in the art can, using well known techniques, isolate the protein that is produced using such expression systems. The methods of purifying proteins from natural sources using antibodies may be equally applied to purifying protein produced by recombinant DNA methodology.

The pharmaceutical compositions of the present invention may be administered by any means that enables the active agent to reach the agent's site of action in the body of a mammal. The pharmaceutical compositions of the present invention may be administered in a number of ways depending upon whether local or systemic treatment is desired and upon the area to be treated. Administration may be topical (including ophthalmic, vaginal, rectal, intranasal, transdermal), oral or parenteral. Parenteral administration includes intravenous drip, subcutaneous, intraperitoneal or intramuscular injection, pulmonary administration, *e.g.*, by inhalation or insufflation, or intrathecal or intraventricular administration.

The present invention further relates to injectable pharmaceutical compositions which comprise such nucleic acid molecules.

The injectable pharmaceutical compositions that comprise a modified nucleotide sequence operably linked to regulatory elements may be delivered using any of several well known technologies including DNA injection (also referred to as DNA vaccination), recombinant vectors such as recombinant adenovirus, recombinant adenovirus associated virus and recombinant vaccinia.

DNA vaccines are described in U.S. Patent Nos. 5,593,972, 5,739,118, 5,817,637, 5,830,876, 5,962,428, 5,981,505, 5,580,859, 5,703,055, 5,676,594, and the priority applications cited therein, which are each incorporated herein by reference. In addition to the delivery protocols described in those applications, alternative methods of delivering DNA are described in U.S. Patent Nos. 4,945,050 and 5,036,006, which are both incorporated herein by reference.

Routes of administration include, but are not limited to, intramuscular, intranasally, intraperitoneal, intradermal, subcutaneous, intravenous, intraarterially, intraoccularly and oral as well as topically, transdermally, by inhalation or suppository or to mucosal tissue such as by lavage to vaginal, rectal, urethral, buccal and sublingual tissue. Preferred routes of administration include to mucosal tissue, intramuscular, intraperitoneal, intradermal and subcutaneous injection. Genetic constructs may be administered by means including, but not limited to, traditional syringes, needleless injection devices, or "microprojectile bombardment gene guns".

When taken up by a cell, the genetic construct(s) may remain present in the cell as a functioning extrachromosomal molecule and/or integrate into the cell's chromosomal DNA. DNA may be introduced into cells where it remains as separate genetic material in the form of a plasmid or plasmids. Alternatively, linear DNA which can integrate into the chromosome may be introduced into the cell. When introducing DNA into the cell, reagents which promote DNA integration into chromosomes may be added. DNA sequences which are useful to promote integration may also be included in the DNA molecule. Alternatively, RNA may be administered to the cell. It is also contemplated to provide the genetic construct as a linear minichromosome including a centromere, telomeres and an origin of replication. Gene constructs may remain part of the genetic material in attenuated live microorganisms or recombinant microbial vectors which live in cells. Gene constructs may be part of genomes of recombinant viral vaccines where the genetic material either integrates into the chromosome of the cell or remains extrachromosomal.

Genetic constructs include regulatory elements necessary for gene expression of a nucleic acid molecule. The elements include: a promoter, an initiation codon, a stop codon, and a polyadenylation signal. In addition, enhancers are often required for gene expression of the sequence that encodes the target protein. It is necessary that these elements be operable linked to the sequence that encodes the desired proteins and that the regulatory elements are operably in the individual to whom they are administered. Initiation codons and stop codon are generally considered to be part of a nucleotide sequence that encodes the desired protein. However, it is necessary that these elements are functional in the individual to whom the gene construct is administered. The initiation and termination codons must be in frame with the coding sequence. Promoters and polyadenylation signals used must be functional within the cells of the individual. Examples of promoters useful to practice the present invention, especially in the production of

a genetic vaccine for humans, include but are not limited to promoters from Simian Virus 40 (SV40), Mouse Mammary Tumor Virus (MMTV) promoter, Human Immunodeficiency Virus (HIV) such as the HIV Long Terminal Repeat (LTR) promoter, Moloney virus, ALV, Cytomegalovirus (CMV) such as the CMV immediate early promoter, Epstein Barr Virus (EBV), Rous Sarcoma Virus (RSV) as well as promoters from human genes such as human Actin, human Myosin, human Hemoglobin, human muscle creatine and human metalothionein. Examples of polyadenylation signals useful to practice the present invention, especially in the production of a genetic vaccine for humans, include but are not limited to SV40 polyadenylation signals and LTR polyadenylation signals. In particular, the SV40 polyadenylation signal which is in pCEP4 plasmid (Invitrogen, San Diego CA), referred to as the SV40 polyadenylation signal, is used. In addition to the regulatory elements required for DNA expression, other elements may also be included in the DNA molecule. Such additional elements include enhancers. The enhancer may be selected from the group including but not limited to: human Actin, human Myosin, human Hemoglobin, human muscle creatine and viral enhancers such as those from CMV, RSV and EBV. Genetic constructs can be provided with mammalian origin of replication in order to maintain the construct extrachromosomally and produce multiple copies of the construct in the cell. Plasmids pCEP4 and pREP4 from Invitrogen (San Diego, CA) contain the Epstein Barr virus origin of replication and nuclear antigen EBNA-1 coding region which produces high copy episomal replication without integration.

One method of the present invention comprises the steps of administering nucleic acid molecules intramuscularly, intranasally, intraperatoneally, subcutaneously, intradermally, or topically or by lavage to mucosal tissue selected from the group consisting of inhalation, vaginal, rectal, urethral, buccal and sublingual.

In some embodiments, the nucleic acid molecule is delivered to the cells in conjunction with administration of a polynucleotide function enhancer or a genetic vaccine facilitator agent. Polynucleotide function enhancers are described in U.S. Serial Number 08/008,342 filed January 26, 1993, U.S. Serial Number 08/029,336 filed March 11, 1993, U.S. Serial Number 08/125,012 filed September 21, 1993, and International Application Serial Number PCT/US94/00899 filed January 26, 1994, which are each incorporated herein by reference. Genetic vaccine facilitator (GVF) agents are described in U.S. Serial Number 08/221,579 filed April 1, 1994, which is incorporated herein by reference. The co-agents which are administered in conjunction with nucleic acid molecules may be administered as a mixture with the nucleic acid molecule or

administered separately simultaneously, before or after administration of nucleic acid molecules. In addition, other agents which may function as transfecting agents and/or replicating agents and/or inflammatory agents and which may be co-administered with a GVF include growth factors, cytokines and lymphokines such as α-interferon, gamma-interferon, platelet derived growth factor (PDGF), TNF, epidermal growth factor (EGF), IL-1, IL-2, IL-4, IL-6, IL-10 and IL-12 as well as fibroblast growth factor, surface active agents such as immune-stimulating complexes (ISCOMS), Freund's incomplete adjuvant, LPS analog including monophosphoryl Lipid A (MPL), muramyl peptides, quinone analogs and vesicles such as squalene and squalene, and hyaluronic acid may also be used administered in conjunction with the genetic construct. In some embodiments, an immunomodulating protein may be used as a GVF.

The pharmaceutical compositions according to the present invention comprise about 1 nanogram to about 2000 micrograms of DNA. In some preferred embodiments, pharmaceutical compositions according to the present invention comprise about 5 nanogram to about 1000 micrograms of DNA. In some preferred embodiments, the pharmaceutical compositions contain about 10 nanograms to about 800 micrograms of DNA. In some preferred embodiments, the pharmaceutical compositions contain about 0.1 to about 500 micrograms of DNA. In some preferred embodiments, the pharmaceutical compositions contain about 1 to about 350 micrograms of DNA. In some preferred embodiments, the pharmaceutical compositions contain about 25 to about 250 micrograms of DNA. In some preferred embodiments, the pharmaceutical compositions contain about 100 to about 200 micrograms DNA.

The pharmaceutical compositions according to the present invention are formulated according to the mode of administration to be used. In cases where pharmaceutical compositions are injectable pharmaceutical compositions, they are sterile, pyrogen free and particulate free. An isotonic formulation is preferably used. Generally, additives for isotonicity can include sodium chloride, dextrose, mannitol, sorbitol and lactose. In some cases, isotonic solutions such as phosphate buffered saline are preferred. Stabilizers include gelatin and albumin. In some embodiments, a vaso-constriction agent is added to the formulation.

The present invention is useful to elicit broad immune responses against a target protein, *i.e.*, proteins specifically associated with pathogens, allergens or the individual's own "abnormal" cells. The present invention is useful to immunize individuals against pathogenic agents and organisms such that an immune response against a pathogen protein provides protective immunity against the pathogen. The present invention is useful to combat hyperproliferative

diseases and disorders such as cancer by eliciting an immune response against a target protein that is specifically associated with the hyperproliferative cells. The present invention is useful to combat autoimmune diseases and disorders by eliciting an immune response against a target protein that is specifically associated with cells involved in the autoimmune condition.

The nucleic acid molecule(s) may be provided as plasmid DNA, the nucleic acid molecules of recombinant vectors or as part of the genetic material provided in an attenuated vaccine or cell vaccine. Alternatively, in some embodiments, the target protein and/or wither or both immunomodulating proteins may be delivered as a protein in addition to the nucleic acid molecules that encode them or instead of the nucleic acid molecules that encode them.

The present invention may be used to immunize an individual against all pathogens such as viruses, prokaryotic and pathogenic eukaryotic organisms such as unicellular pathogenic organisms and multicellular parasites. The present invention is particularly useful to immunize an individual against those pathogens which infect cells and which are not encapsulated such as viruses, and prokaryotes such as gonorrhea, listeria and shigella. In addition, the present invention is also useful to immunize an individual against protozoan pathogens which include a stage in the life cycle where they are intracellular pathogens.

In order to produce a genetic vaccine to protect against pathogen infection, genetic material which encodes immunogenic proteins against which a protective immune response can be mounted must be included in a genetic construct as the coding sequence for the target. Whether the pathogen infects intracellularly, for which the present invention is particularly useful, or extracellularly, it is unlikely that all pathogen antigens will elicit a protective response. Because DNA and RNA are both relatively small and can be produced relatively easily, the present invention provides the additional advantage of allowing for vaccination with multiple pathogen antigens. The genetic construct used in the genetic vaccine can include genetic material which encodes many pathogen antigens. For example, several viral genes may be included in a single construct thereby providing multiple targets.

Another aspect of the present invention provides a method of conferring a broad based protective immune response against hyperproliferating cells that are characteristic in hyperproliferative diseases and to a method of treating individuals suffering from hyperproliferative diseases. Examples of hyperproliferative diseases include all forms of cancer and psoriasis.

It has been discovered that introduction of a genetic construct that includes a nucleotide sequence which encodes an immunogenic "hyperproliferating cell"-associated protein into the cells of an individual results in the production of those proteins in the vaccinated cells of an individual. To immunize against hyperproliferative diseases, a genetic construct that includes a nucleotide sequence which encodes a protein that is associated with a hyperproliferative disease is administered to an individual.

In order for the hyperproliferative-associated protein to be an effective immunogenic target, it must be a protein that is produced exclusively or at higher levels in hyperproliferative cells as compared to normal cells. Target antigens include such proteins, fragments thereof and peptides which comprise at least an epitope found on such proteins. In some cases, a hyperproliferative-associated protein is the product of a mutation of a gene that encodes a protein. The mutated gene encodes a protein which is nearly identical to the normal protein except it has a slightly different amino acid sequence which results in a different epitope not found on the normal protein. Such target proteins include those which are proteins encoded by oncogenes such as *myb, myc, fyn,* and the translocation gene *bcr/abl*, *ras, src*, P53, *neu, trk* and EGRF. In addition to oncogene products as target antigens, target proteins for anti-cancer treatments and protective regimens include variable regions of antibodies made by B cell lymphomas and variable regions of T cell receptors of T cell lymphomas which, in some embodiments, are also used target antigens for autoimmune disease. Other tumor-associated proteins can be used as target proteins such as proteins which are found at higher levels in tumor cells including the protein recognized by monoclonal antibody 17-1A and folate binding proteins.

While the present invention may be used to immunize an individual against one or more of several forms of cancer, the present invention is particularly useful to prophylactically immunize an individual who is predisposed to develop a particular cancer or who has had cancer and is therefore susceptible to a relapse. Developments in genetics and technology as well as epidemiology allow for the determination of probability and risk assessment for the development of cancer in individual. Using genetic screening and/or family health histories, it is possible to predict the probability a particular individual has for developing any one of several types of cancer.

Similarly, those individuals who have already developed cancer and who have been treated to remove the cancer or are otherwise in remission are particularly susceptible to relapse

and reoccurrence. As part of a treatment regimen, such individuals can be immunized against the cancer that they have been diagnosed as having had in order to combat a recurrence. Thus, once it is known that an individual has had a type of cancer and is at risk of a relapse, they can be immunized in order to prepare their immune system to combat any future appearance of the cancer.

The present invention provides a method of treating individuals suffering from hyperproliferative diseases. In such methods, the introduction of genetic constructs serves as an immunotherapeutic, directing and promoting the immune system of the individual to combat hyperproliferative cells that produce the target protein.

The present invention provides a method of treating individuals suffering from autoimmune diseases and disorders by conferring a broad based protective immune response against targets that are associated with autoimmunity including cell receptors and cells which produce "self"-directed antibodies.

T cell mediated autoimmune diseases include rheumatoid arthritis (RA), multiple sclerosis (MS), Sjogren's syndrome, sarcoidosis, insulin dependent diabetes mellitus (IDDM), autoimmune thyroiditis, reactive arthritis, ankylosing spondylitis, scleroderma, polymyositis, dermatomyositis, psoriasis, vasculitis, Wegener's granulomatosis, Crohn's disease, and ulcerative colitis. Each of these diseases is characterized by T cell receptors that bind to endogenous antigens and initiate the inflammatory cascade associated with autoimmune diseases. Vaccination against the variable region of the T cells would elicit an immune response including CTLs to eliminate those T cells.

In RA, several specific variable regions of T cell receptors (TCRs) which are involved in the disease have been characterized. These TCRs include Vβ-3, Vβ-14, Vβ-17 and Vα-17. Thus, vaccination with a DNA construct that encodes at least one of these proteins will elicit an immune response that will target T cells involved in RA. *See*: Howell *et al.*, 1991, Proc. Natl. Acad. Sci. USA, 88:10921-10925; Paliard *et al.*, 1991, Science, 253:325-329; Williams *et al.*, 1992, J. Clin. Invest., 90:326-333; each of which is incorporated herein by reference.

In MS, several specific variable regions of TCRs which are involved in the disease have been characterized. These TCRs include Vβ-7 and Vα-10. Thus, vaccination with a DNA construct that encodes at least one of these proteins will elicit an immune response that will target T cells involved in MS. *See*: Wucherpfennig *et al.*, 1990, Science, 248:1016-1019; Oksenberg *et al.*, 1990, Nature, 345:344-346; each of which is incorporated herein by reference.

In scleroderma, several specific variable regions of TCRs which are involved in the disease have been characterized. These TCRs include Vβ-6, Vβ-8, Vβ-14 and Vα-16, Vα-3C, Vα-7, Vα-14, Vα-15, Vα-16, Vα-28 and Vα-12. Thus, vaccination with a DNA construct that encodes at least one of these proteins will elicit an immune response that will target T cells involved in scleroderma.

In order to treat patients suffering from a T cell mediated autoimmune disease, particularly those for which the variable region of the TCR has yet to be characterized, a synovial biopsy can be performed. Samples of the T cells present can be taken and the variable region of those TCRs identified using standard techniques. Genetic vaccines can be prepared using this information.

B cell mediated autoimmune diseases include systemic lupus erythematosus (SLE), Grave's disease, myasthenia gravis, autoimmune hemolytic anemia, autoimmune thrombocytopenia, asthma, cryoglobulinemia, primary biliary sclerosis, and pernicious anemia. Each of these diseases is characterized by antibodies which bind to endogenous antigens and initiate the inflammatory cascade associated with autoimmune diseases. Vaccination against the variable region of antibodies would elicit an immune response including CTLs to eliminate those B cells that produce the antibody.

In order to treat patients suffering from a B cell mediated autoimmune disease, the variable region of the antibodies involved in the autoimmune activity must be identified. A biopsy can be performed and samples of the antibodies present at a site of inflammation can be taken. The variable region of those antibodies can be identified using standard techniques. Genetic vaccines can be prepared using this information.

In the case of SLE, one antigen is believed to be DNA. Thus, in patients to be immunized against SLE, their sera can be screened for anti-DNA antibodies and a vaccine can be prepared which includes DNA constructs that encode the variable region of such anti-DNA antibodies found in the sera.

Common structural features among the variable regions of both TCRs and antibodies are well known. The DNA sequence encoding a particular TCR or antibody can generally be found following well known methods such as those described in Kabat *et al.*, 1987, *Sequence of Proteins of Immunological Interest*, U.S. Department of Health and Human Services, Bethesda MD, which is incorporated herein by reference. In addition, a general method for cloning

functional variable regions from antibodies can be found in Chaudhary *et al.*, 1990, Proc. Natl. Acad. Sci. USA, 87:1066, which is incorporated herein by reference.

In some of the embodiments of the invention that relate to gene therapy, the gene constructs contain either compensating genes or genes that encode therapeutic proteins. Examples of compensating genes include a gene which encodes dystrophin or a functional fragment, a gene to compensate for the defective gene in patients suffering from cystic fibrosis, an insulin, a gene to compensate for the defective gene in patients suffering from ADA, and a gene encoding Factor VIII. Examples of genes encoding therapeutic proteins include genes which encodes erythropoietin, interferon, LDL receptor, GM-CSF, IL-2, IL-4 and TNF. Additionally, genetic constructs which encode single chain antibody components which specifically bind to toxic substances can be administered. In some preferred embodiments, the dystrophin gene is provided as part of a mini-gene and used to treat individuals suffering from muscular dystrophy. In some preferred embodiments, a mini-gene which contains coding sequence for a partial dystrophin protein is provided. Dystrophin abnormalities are responsible for both the milder Becker's Muscular Dystrophy (BMD) and the severe Duchenne's Muscular Dystrophy (DMD). In BMD dystrophin is made, but it is abnormal in either size and/or amount. The patient is mild to moderately weak. In DMD no protein is made and the patient is chair-bound by age 13 and usually dies by age 20. In some patients, particularly those suffering from BMD, partial dystrophin protein produced by expression of a mini-gene delivered according to the present invention can provide improved muscle function.

In some preferred embodiments, genes encoding IL-2, IL-4, interferon, or TNF are delivered to tumor cells which are either present or removed and then reintroduced into an individual. In some embodiments, a gene encoding gamma interferon is administered to an individual suffering from multiple sclerosis.

In addition to using modified nucleic acid sequences to improve genetic vaccines, the present invention relates to improved attenuated live vaccines and improved vaccines which use recombinant vectors to deliver foreign genes that encode antigens. Examples of attenuated live vaccines and those using recombinant vectors to deliver foreign antigens are described in U.S. Patent Nos.: 4,722,848; 5,017,487; 5,077,044; 5,110,587; 5,112,749; 5,174,993; 5,223,424; 5,225,336; 5,240,703; 5,242,829; 5,294,441; 5,294,548; 5,310,668; 5,387,744; 5,389,368; 5,424,065; 5,451,499; 5,453,364; 5,462,734; 5,470,734; and 5,482,713, which are each incorporated herein by reference. Gene constructs are provided which include the modified

nucleotide sequence operably linked to regulatory sequences that can function in the vaccinee to effect expression. The gene constructs are incorporated in the attenuated live vaccines and recombinant vaccines to produce improved vaccines according to the invention. Likewise modified nucleic acid sequences can be used in recombinant vectors useful to deliver gene therapeutics that encode desired proteins.

The present invention provides an improved method of immunizing individuals that comprises the step of delivering gene constructs to the cells of individuals as part of vaccine compositions which include are provided which include DNA vaccines, attenuated live vaccines and recombinant vaccines. The gene constructs comprise a nucleotide sequence that encodes an immunomodulating protein and that is operably linked to regulatory sequences that can function in the vaccinee to effect expression. The improved vaccines result in an enhanced cellular immune response.

The invention is further illustrated by way of the following examples, which are intended to elaborate several embodiments of the invention. These examples are not intended, nor are they to be construed, as limiting the scope of the invention. It will be clear that the invention may be practiced otherwise than as particularly described herein. Numerous modifications and variations of the present invention are possible in view of the teachings herein and, therefore, are within the scope of the invention.

EXAMPLES

Example 1: Materials and Methods.

Prediction of mRNA secondary structure

To enhance translation efficiency of transgenes, RNA secondary structure was predicted by using MulFold and viewed by LoopDloop software for the Macintosh computer.

Immunoprecipitation of radiolabeled *in vitro* translated proteins

^{35}S-labeled protein products were prepared using the TNT-T7 coupled Transcription/Translation System (Promega). 10 ml of radiolabeled protein sample and 1 ml of anti-His (C-term) antibody (Invitrogen, CA) were added to 300 µl of RIPA buffer and mixed gently. After an incubation at 4°C for 90 minutes, Protein A-Sepharose beads (Amersham-Pharmacia Biotech, Piscataway, NJ) was added to the protein-antibody complexes at a final concentration of 5 mg per tube and the samples were then incubated at 4°C for 90 minutes in a rotating shaker. The beads were washed three times with RIPA buffer and suspended in 2X SDS

sample buffer. The immunoprecipitated protein complexes were eluted from the Sepharose beads by brief boiling and resolved in SDS/PAGE (15%) gels. The mobility of the protein samples was compared with that of commercially available ^{14}C-methylated molecular weight marker (Sigma-Aldrich Corp., St. Louis, MO). The gel was fixed, treated briefly with 1M sodium salicylate solution and dried in a gel drier (BioRad, Hercules, CA). The dried gel was exposed overnight to X-ray film (Kodak, Rochester, NY). The molecular size of the *in vitro* translated protein was 21.5kD.

***In vitro* translated protein**

Non-radioactive, *in vitro* translated Cp protein was also generated as described above, using the TNT-T7 coupled Transcription/Translation System (Promega, Madison, WI) with non-radioactive components. An *in vitro* translation control was generated using the *in vitro* translation kit with the pcDNA3.1 vector (Invitrogen, San Diego, CA), lacking an expressible insert.

DNA inoculation of mice

The quadriceps muscles of 6- to 8-week-old female BALB/c mice (Harlan Sprague Dawley, Inc., Indianapolis, IN) were injected with 100 μg of pWNVh-DJY, pWNVy-DJY, or pcDNA3.1 in phosphate buffered saline (PBS) and 0.25% bupivacaine-HCl (Sigma, St. Louis, MO). Mice were injected with two DNA immunizations (100 μg each) separated by two weeks. At thirteen days after the boost injection, the mice were sacrificed, the spleens were harvested, and the lymphocytes were isolated and tested for cellular immune responses.

Intracellular IFN-γ detection by flow cytometry

In each well of a round-bottom 96-well plate was placed 100 μl of RPMI-1640 (supplemented with 5 % FBS), containing 50 U/ml rHuIL-2 (Intergen, Purchase, NY), 10 μg/ml Brefeldin A (Pharmingen, San Diego, CA), 100 ng/ml PMA (Sigma, St. Louis, MO), and 1 μg/ml ionomycin (Sigma, St. Louis, MO). Either *in vitro* translated protein or an *in vitro* translation control (generated using the *in vitro* translation kit with the vector backbone lacking an expressible insert), at 4 μg/ml was added in 50 μl of R5 medium. After adding the antigens (Ags), isolated splenocytes were added to each well at $1x10^6$ cells in 50 μl of R5 medium. For the compensation in flow cytometry, splenocytes from naïve mice were set up with only IL-2 and Brefeldin A. The plates were incubated in 37 °C, 5 % CO_2 in an incubator for 5 to 6 hours. As a control, cells were incubated without Ag. After incubation, the plate was spun at 1200 rpm for 5 minutes and the supernatants discarded. The cells were resuspended with 200 μl of PBS,

supplemented with 1% BSA, and put on ice for 15 minutes, and then spun down and resuspended with anti-CD4-PE mAb (Pharmingen) at 0.1 μg/sample in 50 μl of PBS/1% BSA. After incubation for 30 minutes at 4°C, the cells were washed twice with PBS/1%. After the second wash, cell pellets were resuspended with 100 μl of Cytofix/Cytoperm solution (Pharmingen) and incubated for 20 minutes at 4°C. The cells were washed twice with 1 x Perm/Wash (Pharmingen) and resuspended with 50 μl of Perm/Wash solution containing anti-IFN-γ-APC (Pharmingen) at 0.1 μg/sample concentration. After incubation for 30 minutes at 4°C, the cells were washed twice with 1x Perm/Wash solution and fixed with 2% paraformaldehyde, and then stored at 4°C until analyzed by flow cytometry.

Example 2: Addition of Leader Sequence to West Nile Virus Capsid mRNA.

The addition of a leader sequence to minimize free energy in the West Nile Virus Capsid mRNA resulted in enhanced protein expression and immune response.

To enhance the transcription and translation efficiency of transgenes, the human IgE leader sequence was added to the 5' upstream of open reading frame (orf) sequences (Fig.1).

The addition of a sequence encoding the human IgE leader sequence containing codons that are less prevalently utilized in humans (WNVy-DJY construct (yeast codon)) resulted in a predicted secondary structure for the mRNA having an increased free energy value, relative to the secondary structure for the mRNA without the leader sequence (WNVwt construct (wild type)), or relative to the secondary structure for the mRNA encoding a leader sequence optimized with human codons (WNVh-DJY construct (human codon)) (Fig. 2).

Furthermore, the construct encoding the leader sequence containing codons that are less prevalently utilized in humans (yeast optimized) yielded a higher level of protein than did the construct encoding the leader sequence containing human optimized codons, as determined by immunoprecipitation of radiolabeled *in vitro* translated proteins (Fig. 3; Table 1, yeast codon usage). The codons more prevalently used by yeast are, in general, AU rich; the codons more prevalently used by *Homo sapiens* are, in general, more GC rich (*see* Kim *et al.*, 1997, Gene, *supra*).

Table 1. Yeast codon prevalent usage.

Amino Acid		Yeast codon
A	Ala	GCU
R	Arg	AGA
N	Asn	AAU
D	Asp	GAU
C	Cys	UGU
Q	Gln	CAA
E	Glu	GAA
G	Gly	GGU
H	His	CAU
I	Ile	AUU
L	Leu	UUA
K	Lys	AAA
P	Pro	CCA
F	Phe	UUU
S	Ser	UCU
T	Thr	ACU
W	Trp	UGG
Y	Tyr	UAU
V	Val	GUU

DNA plasmid injection into mouse muscle induced an antigen-specific, $CD4^+$ Th cell-dependent immune response, as determined by intracellular IFN-γ/flow cytometry analysis. The $CD4^+$ Th cell-dependent, intracellular IFN-γ production was quantitated by flow cytometry. Splenocytes isolated from pWNVy-DJY (pWNVCy)-immunized mice, expressed higher levels of IFN-γ upon stimulation with *in vitro* translated Cp protein, than did the splenocytes isolated from pWNVh-DJY (pWNVCh)-immunized mice (see Fig. 4)

Example 3: Removal of RNA Secondary Structure in HIV-1 *pol* RNA by Increasing the Minimum Predicted Free Energy.

The strategy of adding a leader encoding sequence and altering the codons to be yeast optimized (less frequently used in human) was applied to the HIV-1 *pol* gene. When nucleic acid sequence encoding the IgE leader sequence with codons less prevalently used in humans (yeast optimized) was added to the 5' end of HIV-1 *pol* gene, the predicted free energy of the energy minimized transcript was increased (Fig. 5).

In HIV-1 *pol* structural gene, several regions of stable secondary structure, located between nucleotide (nt) 1738 and nt 1938, were predicted by MulFold analysis (Fig. 6). Alteration of the codons in the region from nt 1738 to nt 1938 to codons less prevalently utilized

in humans (yeast optimized codons) resulted in a weakening of the secondary structure in that region. The predicted secondary structure for the region with the modified codons had a higher free energy than the predicted secondary structure for the original sequence (Fig. 7). In addition, the formation of mRNA secondary structure in the first 200 nucleotides of the *pol* gene was minimized by using codons less prevalently utilized in humans (yeast optimized codons) (HIV-1 Pol yt), as compared to a transcript containing the most prevalently utilized codons in humans (human optimized codons) (HIV-1 Pol hu) (Fig. 8). The minimum free energy was dramatically increased from –53.0 kcal to –26.4 kcal.

Example 4: Removal of RNA Secondary Structure in HIV-1 *gag* RNA by Increasing the Minimum Predicted Free Energy.

Several regions of regions of stable secondary structure were predicted by MulFold analysis for the transcript for the HIV-1 *gag* structural gene (Fig. 9), and the minimum free energy was increased (from –351.07 kcal to –283.11 kcal) by using codons that are utilized less prevalently in humans (yeast optimized) (Fig. 10).

Example 5: Removal of RNA Secondary Structure in WNV *env* RNA by Increasing the Minimum Predicted Free Energy.

In the West Nile Virus envelope (*env*) gene, application of the strategy of mRNA energy minimization in the first 200 base pairs (bp) of the gene with codons that are utilized less prevalently in humans (yeast optimized, WNVyt200) increased the minimum free energy of the cognate transcript as compared to the transcript for the wild type WNV *env* gene (WNVwt200) or as compared to a transcript optimized with the most prevalently used codons in humans (WNVhu200) (Fig. 11).

The foregoing examples are meant to illustrate the invention and are not to be construed to limit the invention in any way. Those skilled in the art will recognize modifications that are within the spirit and scope of the invention.

All references cited herein are hereby incorporated by reference in their entirety.

What is claimed is:

1. A method of producing a protein in a recombinant expression system that comprises translation of mRNA transcribed from a heterologous DNA sequence in the expression system, said method comprising the steps of:

a) predicting the secondary structure of mRNA transcribed from a native heterologous DNA sequence;

b) modifying the native heterologous DNA sequence to produce a modified heterologous DNA sequence wherein mRNA transcribed from the modified heterologous DNA sequence has a secondary structure having increased free energy compared to that of the secondary structure of the mRNA transcribed from the native heterologous DNA sequence; and

c) using the modified heterologous DNA sequence in the recombinant expression system for protein production.

2. The method of claim 1, wherein the recombinant expression system is selected from the group consisting of: a cell free *in vitro* transcription and translation system; an *in vitro* cell expression system; a DNA construct used in direct DNA injection; and a recombinant vector for delivery of DNA to an individual.

3. The method of claim 1, wherein the secondary structure of the mRNA transcribed from a native heterologous DNA sequence is predicted using a computer and computer program.

4. The method of claim 1, wherein the secondary structure of the mRNA transcribed from a native heterologous DNA sequence is modified by increasing the AT content of the coding sequence.

5. The method of claim 4, wherein the secondary structure of the mRNA transcribed from a native heterologous DNA sequence is modified by increasing the AT content of the coding sequence at the 5' end of the coding sequence such that mRNA transcribed therefrom has an increased AU content.

6. The method of claim 5, wherein the secondary structure of the mRNA transcribed from a native heterologous DNA sequence is modified by increasing the AT content of the coding

sequence at the 5' end of the coding sequence within 200 nucleotides from the initiation codon such that mRNA transcribed therefrom has an increased AU content.

7. The method of claim 6 wherein the secondary structure of the mRNA transcribed from a native heterologous DNA sequence is modified by increasing the AT content of the coding sequence at the 5' end of the coding sequence within 150 nucleotides from the initiation codon such that mRNA transcribed therefrom has an increased AU content.

8. The method of claim 6 wherein the secondary structure of the mRNA transcribed from a native heterologous DNA sequence is modified by increasing the AT content of the coding sequence at the 5' end of the coding sequence within 100 nucleotides from the initiation codon such that mRNA transcribed therefrom has an increased AU content.

9. An injectable pharmaceutical composition comprising a nucleic acid molecule that includes a modified coding sequence encoding a protein operably linked to regulatory elements, wherein the modified coding sequence comprises a higher AT or AU content relative to the AT or AU content of the native coding sequence, and further comprising a pharmaceutically acceptable carrier.

10. The injectable pharmaceutical composition of claim 9, wherein said modified coding sequence comprises a higher AT or AU content in the first 200 bases relative to the AT or AU content of the native nucleic acid sequence.

11. The injectable pharmaceutical composition of claim 9, wherein said modified coding sequence comprises a higher AT or AU content in the first 150 bases relative to the AT or AU content of the native nucleic acid sequence.

12. The injectable pharmaceutical composition of claim 9, wherein said modified coding sequence comprises a higher AT or AU content in the first 100 bases relative to the AT or AU content of the native nucleic acid sequence.

13. The injectable pharmaceutical composition of claim 9, wherein said modified coding

sequence comprises a higher AT or AU content in at least one region of up to 200 bases in length relative to the AT or AU content of the native nucleic acid sequence.

14. The injectable pharmaceutical composition of claim 9, wherein said modified coding sequence comprises a higher AT or AU content in at least one region of up to 150 bases in length relative to the AT or AU content of the native nucleic acid sequence.

15. The injectable pharmaceutical composition of claim 9, wherein said modified coding sequence comprises a higher AT or AU content in at least one region of up to 100 bases in length relative to the AT or AU content of the native nucleic acid sequence.

16. The injectable pharmaceutical composition of claim 9, wherein the modified coding sequence encodes an immunogen.

17. The injectable pharmaceutical composition of claim 16, wherein the immunogen is a pathogen derived proteins or immunogenic fragment thereof.

18. The injectable pharmaceutical composition of claim 16, wherein the immunogen is a fusion protein that includes a pathogen derived protein or immunogenic fragment thereof.

19. The injectable pharmaceutical composition of claim 16, wherein the immunogen is a cancer antigen or immunogenic fragment thereof.

20. The injectable pharmaceutical composition of claim 16, wherein the immunogen is a fusion protein that includes a cancer antigen or immunogenic fragment thereof.

21. The injectable pharmaceutical composition of claim 16, wherein the immunogen is an autoimmune disease associated protein or immunogenic fragment thereof.

22. The injectable pharmaceutical composition of claim 16, wherein the immunogen is a fusion protein that includes an autoimmune disease associated protein or immunogenic fragment thereof.

23. The injectable pharmaceutical composition of claim 9, wherein the modified coding sequence encodes a non-immunogenic therapeutic protein.

24. The injectable pharmaceutical composition of claim 23, wherein the non-immunogenic therapeutic protein is selected from the group consisting of cytokines, growth factors, blood products, and enzymes.

25. The injectable pharmaceutical composition of claim 9, wherein the modified coding sequence comprises dispersed modifications.

26. The injectable pharmaceutical composition of claim 25, wherein the dispersed modifications are at least two modified coding sequences of up to 200 bases in length alternating with regions of native coding sequence.

27. The injectable pharmaceutical composition of claim 25, wherein the dispersed modifications are at least two modified coding sequences of up to 150 bases in length alternating with regions of native coding sequence.

28. The injectable pharmaceutical composition of claim 25, wherein the dispersed modifications are at least two modified coding sequences of up to 100 bases in length alternating with regions of native coding sequence.

29. The injectable pharmaceutical composition of claim 9, wherein said modified coding sequence comprises a higher AT or AU content in the last 200 bases relative to the AT or AU content of the native nucleic acid sequence.

30. The injectable pharmaceutical composition of claim 9, wherein said modified coding sequence comprises a higher AT or AU content in the last 150 bases relative to the AT or AU content of the native nucleic acid sequence.

31. The injectable pharmaceutical composition of claim 9, wherein said modified coding sequence comprises a higher AT or AU content in the last 100 bases relative to the AT or AU

content of the native nucleic acid sequence.

32. A recombinant viral vector comprising a nucleic acid molecule that includes a modified coding sequence encoding a protein operably linked to regulatory elements, wherein the modified coding sequence comprises a higher AT or AU content relative to the AT or AU content of the native coding sequence.

33. The recombinant viral vector of claim 32, wherein said modified coding sequence comprises a higher AT or AU content in the first 200 bases relative to the AT or AU content of the native nucleic acid sequence.

34. The recombinant viral vector of claim 32, wherein said modified coding sequence comprises a higher AT or AU content in the first 150 bases relative to the AT or AU content of the native nucleic acid sequence.

35. The recombinant viral vector of claim 32, wherein said modified coding sequence comprises a higher AT or AU content in the first 100 bases relative to the AT or AU content of the native nucleic acid sequence.

36. The recombinant viral vector of claim 32, wherein said modified coding sequence comprises a higher AT or AU content in at least one region of up to 200 bases in length relative to the AT or AU content of the native nucleic acid sequence.

37. The recombinant viral vector of claim 32, wherein said modified coding sequence comprises a higher AT or AU content in at least one region of up to 150 bases in length relative to the AT or AU content of the native nucleic acid sequence.

38. The recombinant viral vector of claim 32, wherein said modified coding sequence comprises a higher AT or AU content in at least one region of up to 100 bases in length relative to the AT or AU content of the native nucleic acid sequence.

39. The recombinant viral vector of claim 32, wherein the modified coding sequence encodes

an immunogen.

40. The recombinant viral vector of claim 39, wherein the immunogen is a pathogen derived proteins or immunogenic fragment thereof.

41. The recombinant viral vector of claim 39, wherein the immunogen is a fusion protein that includes a pathogen derived protein or immunogenic fragment thereof.

42. The recombinant viral vector of claim 39, wherein the immunogen is a cancer antigen or immunogenic fragment thereof.

43. The recombinant viral vector of claim 32, wherein the immunogen is a fusion protein that includes a cancer antigen or immunogenic fragment thereof.

44. The recombinant viral vector of claim 39, wherein the immunogen is an autoimmune disease associated protein or immunogenic fragment thereof.

45. The recombinant viral vector of claim 39, wherein the immunogen is a fusion protein that includes an autoimmune disease associated protein or immunogenic fragment thereof.

46. The recombinant viral vector of claim 32, wherein the modified coding sequence encodes a non-immunogenic therapeutic protein.

47. The recombinant viral vector of claim 46, wherein the non-immunogenic therapeutic protein is selected from the group consisting of cytokines, growth factors, blood products, and enzymes.

48. The recombinant viral vector of claim 32, wherein the modified coding sequence comprises dispersed modifications.

49. The recombinant viral vector of claim 48, wherein the dispersed modifications are at least two modified coding sequences of 200 bases in length alternating with regions of native coding

sequence.

50. The recombinant viral vector of claim 48, wherein the dispersed modifications are at least two modified coding sequences of 150 bases in length alternating with regions of native coding sequence.

51. The recombinant viral vector of claim 48, wherein the dispersed modifications are at least two modified coding sequences of 100 bases in length alternating with regions of native coding sequence.

52. The recombinant viral vector of claim 32, wherein said modified coding sequence comprises a higher AT or AU content in the last 200 bases relative to the AT or AU content of the native nucleic acid sequence.

53. The recombinant viral vector of claim 32, wherein said modified coding sequence comprises a higher AT or AU content in the last 150 bases relative to the AT or AU content of the native nucleic acid sequence.

54. The recombinant viral vector of claim 32, wherein said modified coding sequence comprises a higher AT or AU content in the last 100 bases relative to the AT or AU content of the native nucleic acid sequence.

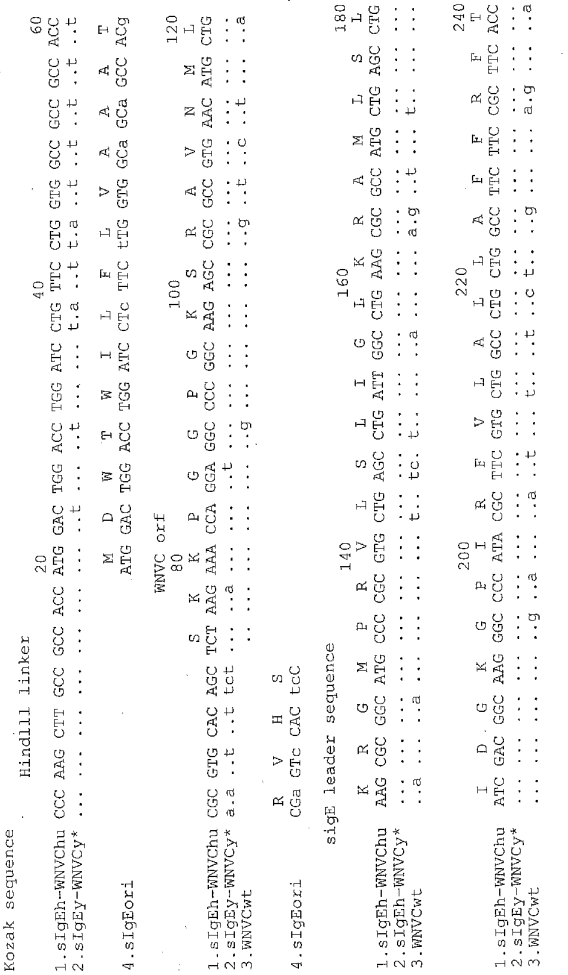

Kozak sequence
HindIII linker
sigE leader sequence
WNVC orf
1.sIgEh-WNVChu
2.sIgEy-WNVCy*
3.WNVCwt
4.sIgEori

FIG. 1a

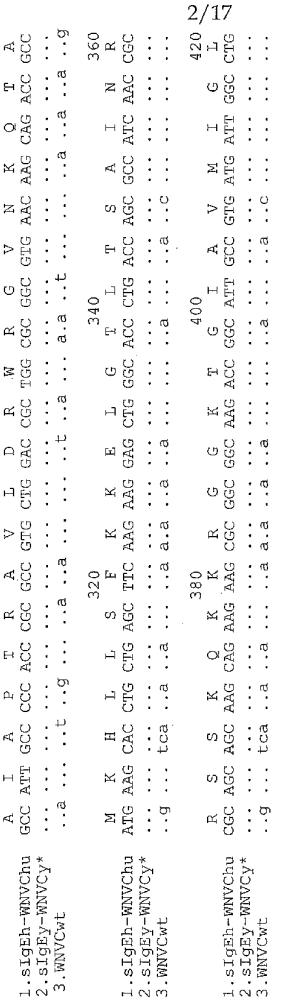

260 280 300
A I A P T R A V L D R W G V N K Q T A
1.sIgEh-WNVChu GCC ATT GCC CCC ACC CGC GCC GTG CTG GAC CGC TGG CGC GGC GTG AAC AAG CAG ACC GCC
2.sIgEy-WNVCy* ... ...
3.WNVCwt ..a ... ..t ..g ... ..a ..a ... .t ..a ... a.a ..t ... .a ..a ..a ..g
320 340 360
M K H L L S F K K E L G T L T S A I N R
1.sIgEh-WNVChu ATG AAG CAC CTG CTG AGC TTC AAG AAG GAG CTG GGC ACC CTG ACC AGC GCC ATC AAC CGC
2.sIgEy-WNVCy* ... ...
3.WNVCwt ..g ... tca ..a ..a ... ..a a.a ..a ..a ... .a ... ..a ..c ... ...
380 400 420
R S S K Q K R G K T G I A V M I G L
1.sIgEh-WNVChu CGC AGC AGC AAG CAG AAG AAG CGC GGC AAG ACC GGC ATT GCC GTG ATG ATT GGC CTG
2.sIgEy-WNVCy* ... ...
3.WNVCwt ..g ... tca ..a ... ..a a.a ..a ..a ... .a ... ..a ..c ... ...
NotI linker
I A S V G A
1.sIgEh-WNVChu ATC GCC AGC GTG GGC GCG GCC GCT AAA CTA T
2.sIgEy-WNVCy* ... ...
3.WNVCwt ... ...a ..a ..a

FIG. 1b

mRNA Energy Minimization in 1st 73 Bases

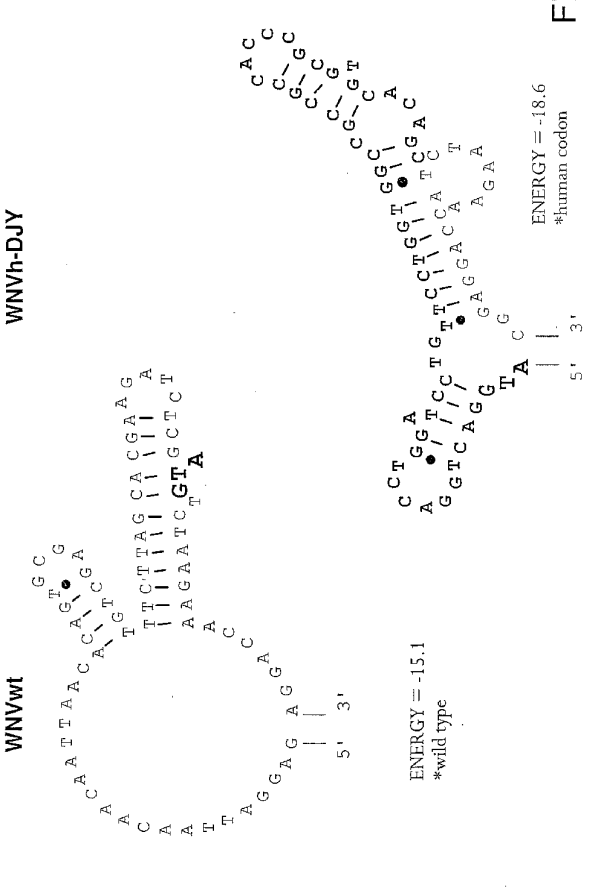
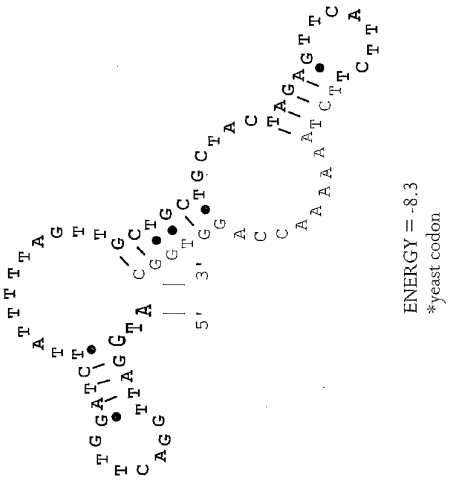


FIG. 2a

FIG. 2b

mRNA Energy Minimization in 1st 73 Bases

WNVy-DJY



Radioimmunoprecipitation with *In Vitro* Translated Proteins

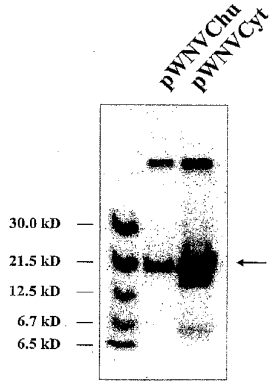


FIG. 3

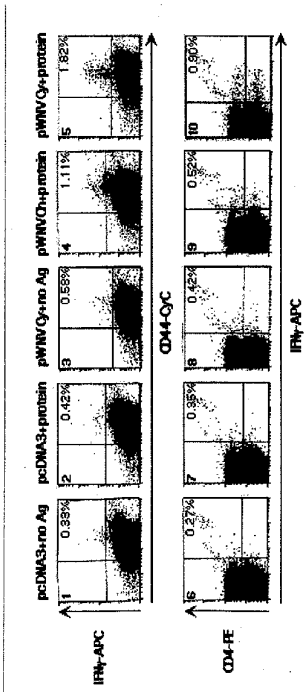

pcDNA3+no Ag
pcDNA3+protein
pWNVGs+no Ag
pWNVCh+protein
pWNVGs+protein
0.33%
0.42%
0.58%
1.11%
1.82%
0.27%
0.35%
0.42%
0.52%
0.90%
CD44-CyC
IFNγ-APC
CD4-PE

FIG. 4

mRNA Energy Minimization: Addition of Leader Sequence

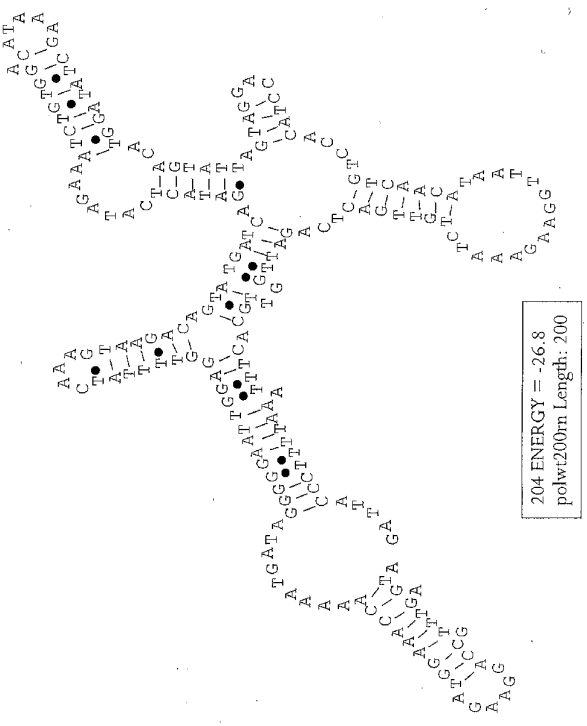


FIG. 5a

mRNA Energy Minimization: Addition of Leader Sequence

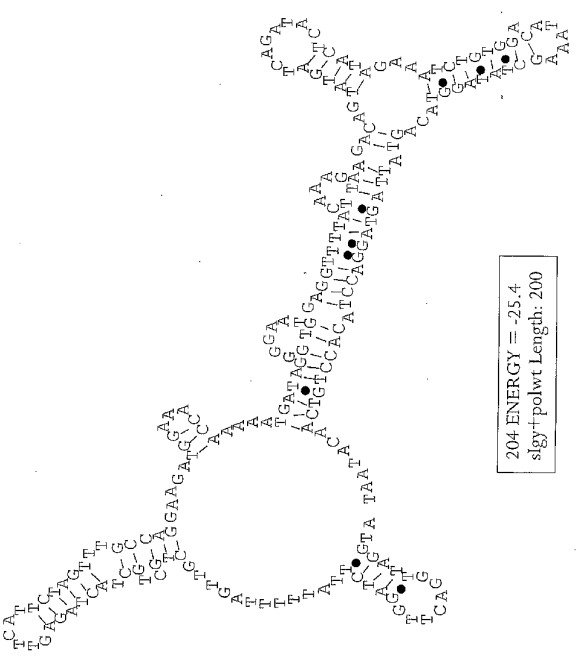
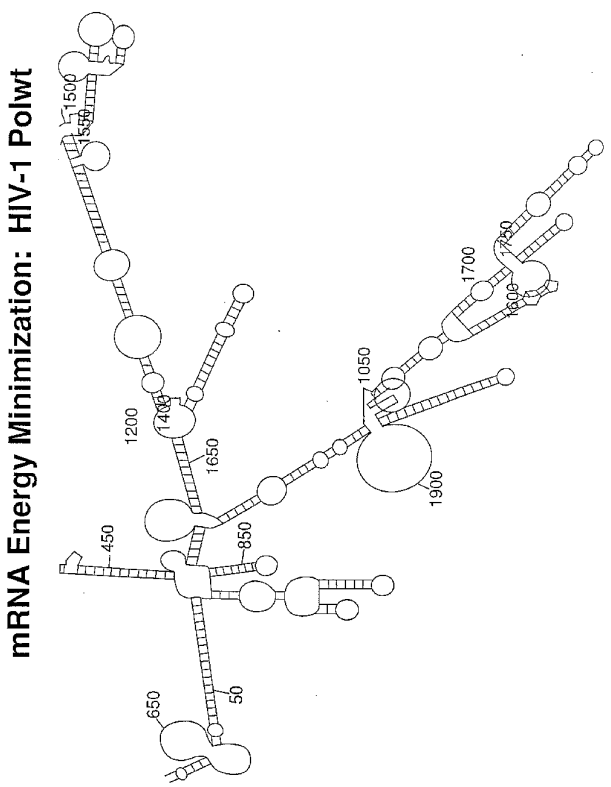


FIG. 5b

FIG. 6


mRNA Energy Minimization: HIV-1 Polwt
1500
1550
1700
1750
1800
1050
1200
1400
1650
1900
450
850
650
50

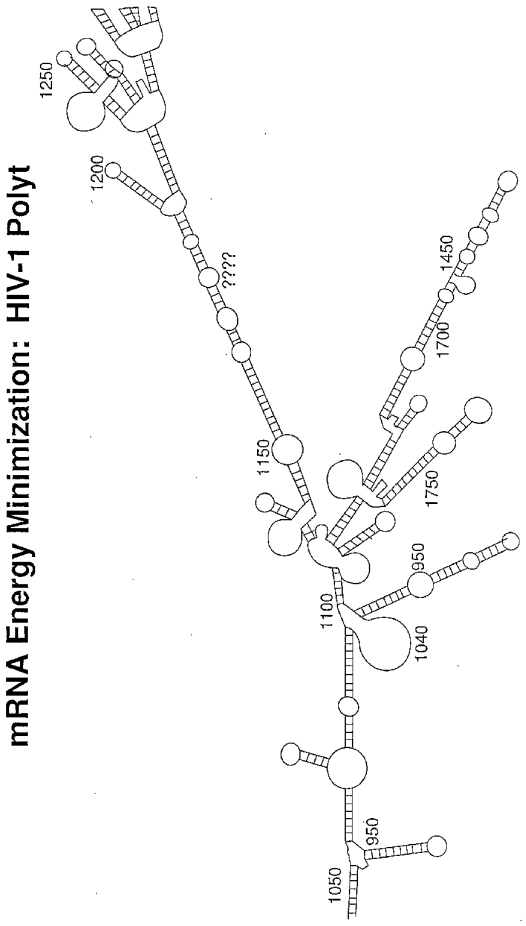
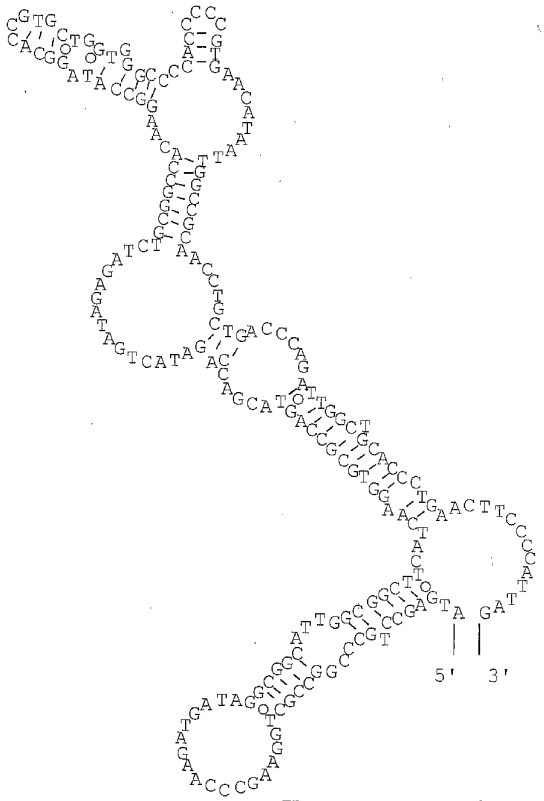

mRNA Energy Minimization: HIV-1 Polyt
1250
1200
????
1150
1450
1700
1750
950
1100
1040
950
1050

FIG. 7

mRNA Energy Minimization in 1st 73 Bases

HIV-1 Pol hu



FREE ENERGY = -53.0 kcal
HIV-1 Polhu Lenght: 200 bases

FIG. 8a

mRNA Energy Minimization in 1st 73 Bases

HIV-1 Pol yt

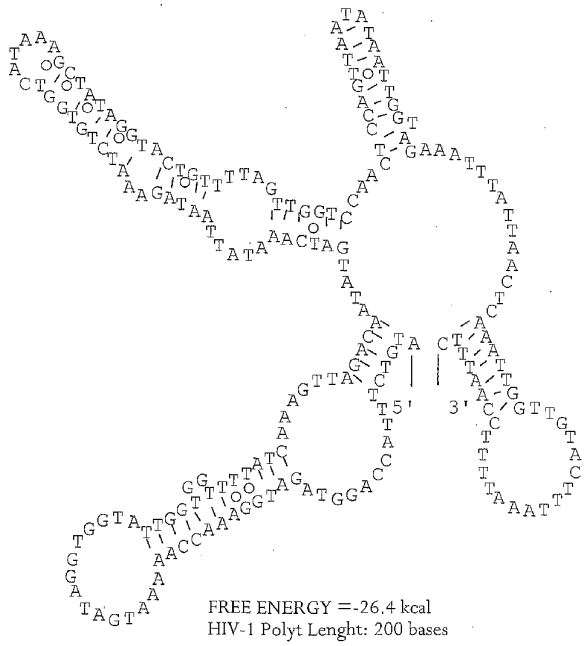


FIG. 8b

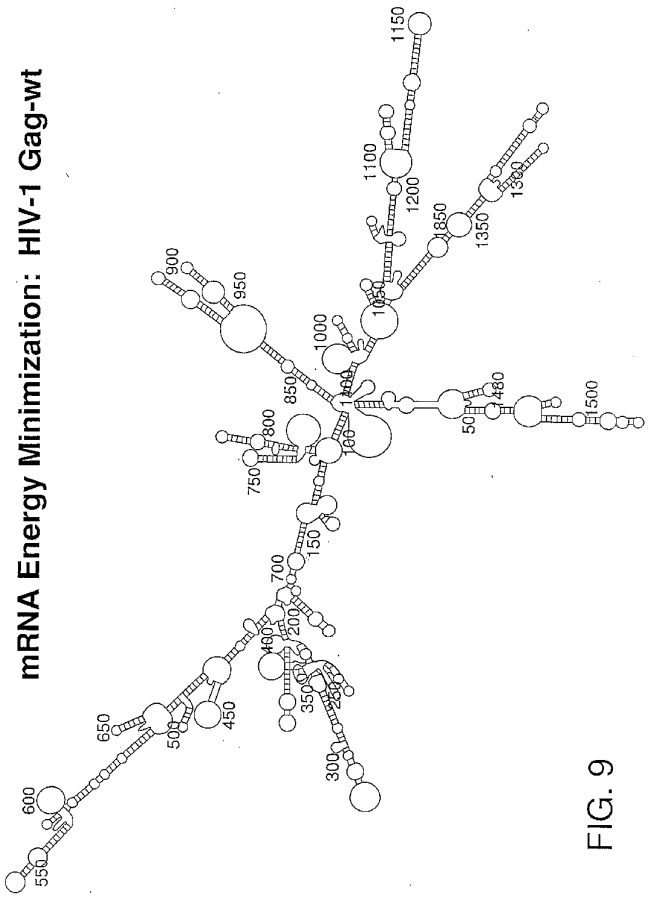

mRNA Energy Minimization: HIV-1 Gag-wt
1150
1100
1200
1350
1300
900
950
1000
850
800
750
700
150
200
400
350
250
300
450
500
650
600
550
1500
1450

FIG. 9

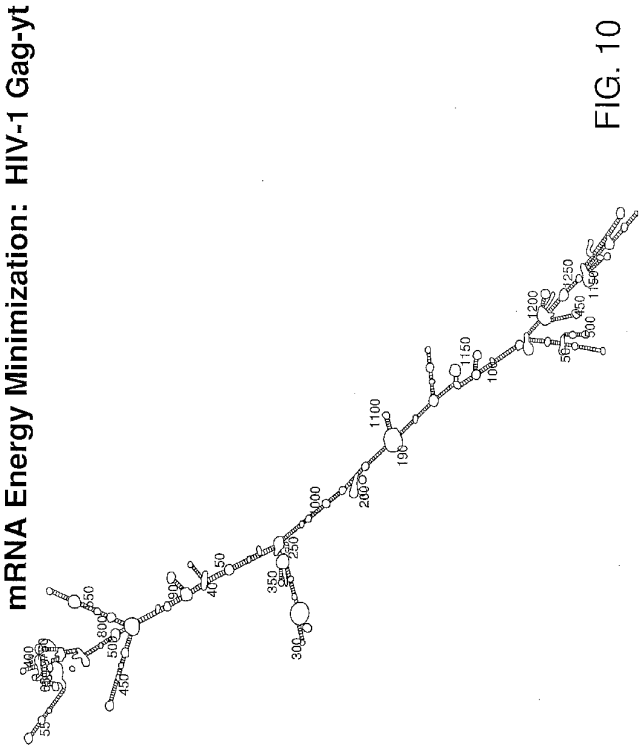


FIG. 10

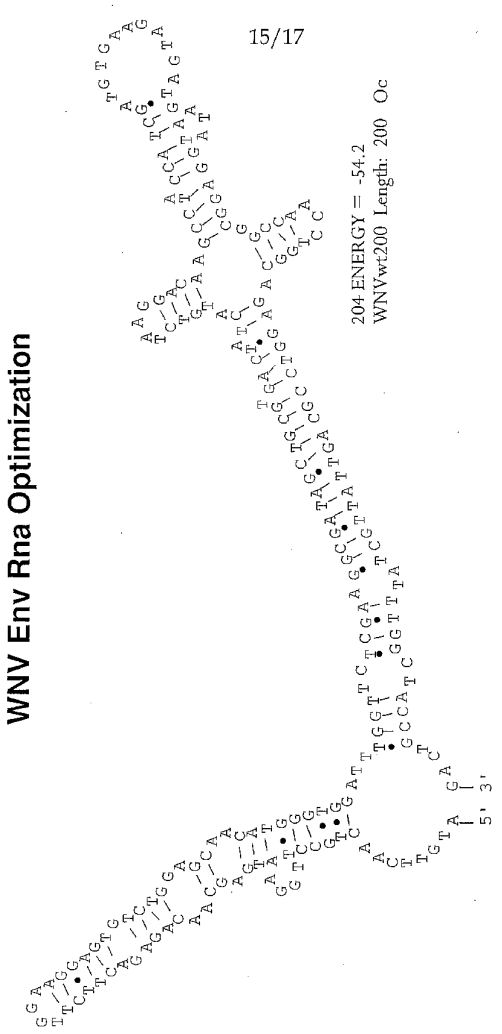

WNV Env Rna Optimization
204 ENERGY = -54.2
WNVwt200 Length: 200 Oc
5'
3'

FIG. 11a

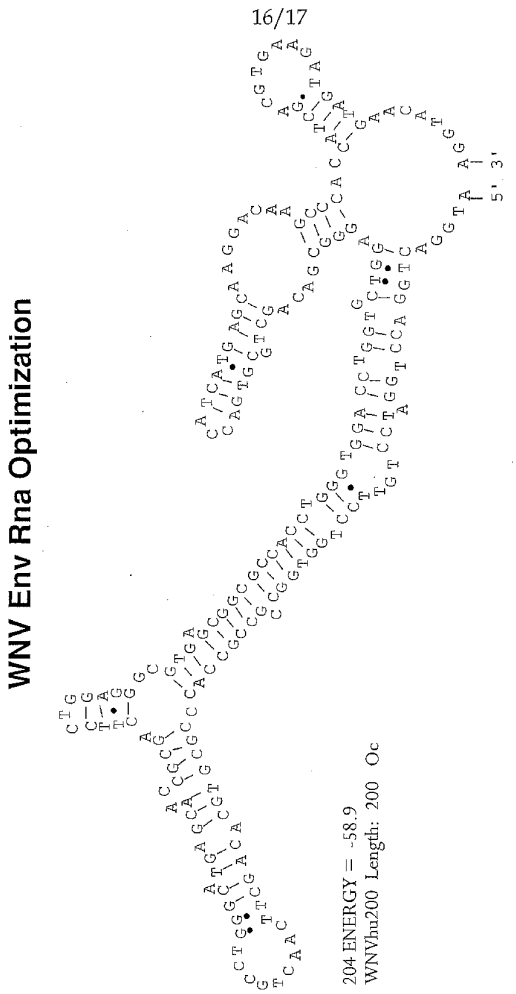

WNV Env Rna Optimization
204 ENERGY = -58.9
WNVhu200 Length: 200 Oc
5' 3'

FIG. 11b

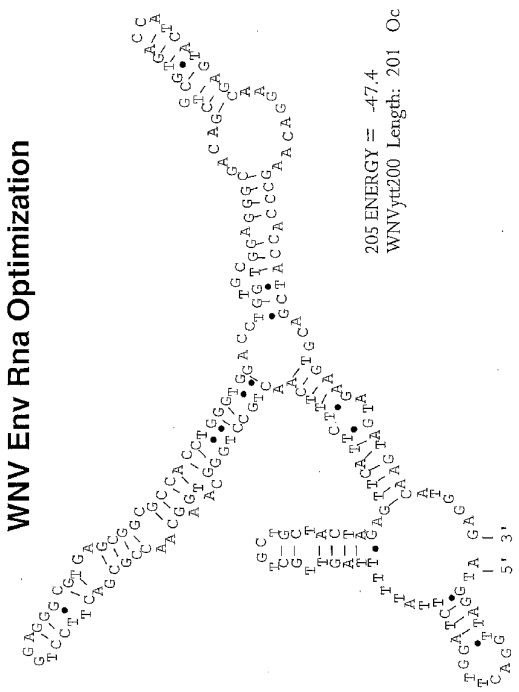

WNV Env Rna Optimization
205 ENERGY = -47.4
WNVytt200 Length: 201 Oc
5' 3'

FIG. 11c

【国際調査報告】

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US01/31451

A. CLASSIFICATION OF SUBJECT MATTER
IPC(7) : C12P 21/00, 21/02; A01N 43/04; C12N 15/00, 15/09, 15/11, 15/63
US CL : 435/69.1, 320.1; 514/44; 536/23.1
According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED
Minimum documentation searched (classification system followed by classification symbols)
U.S. : 435/69.1, 320.1; 514/44; 536/23.1

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
Please See Continuation Sheet

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X --- Y	de SMIT et al, Control of translation by mRNA secondary structure in Escherichia coli, Journal of Molecular Biology, 1994, Vol. 244, pages 144-150, see entire document.	1-8 --- 9-54
X --- Y	MUKUND et al, Effect of mRNA secondary structure in the regulation of gene expression: Unfolding of stable loop causes the expression of Taq polymerase in E. coli, Current Science, 10 June 1999, Vol. 76, No. 11, pages 1486-1490, see entire document.	1-8 --- 9-54
X --- Y	SUO et al, RNA secondary structure switching during DNA synthesis catalyzed by HIV-1 reverse transcriptase, Biochemistry, 1997, Vol. 36, pages 14778-14785, see entire document.	1-8 --- 9-54
X --- Y	US 5,489,529 A (de BOER et al) 6 February 1996 (06.02.1996), see entire document.	1-8 --- 9-54

☒ Further documents are listed in the continuation of Box C. ☐ See patent family annex.

* Special categories of cited documents:
"A" document defining the general state of the art which is not considered to be of particular relevance
"E" earlier application or patent published on or after the international filing date
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
"O" document referring to an oral disclosure, use, exhibition or other means
"P" document published prior to the international filing date but later than the priority date claimed
"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"&" document member of the same patent family

Date of the actual completion of the international search
22 March 2002 (22.03.2002)

Date of mailing of the international search report
24 APR 2002

Name and mailing address of the ISA/US
Commissioner of Patents and Trademarks
Box PCT
Washington, D.C. 20231
Facsimile No. (703)305-3230

Authorized officer
James Martinell
Telephone No. (703) 308-0196

Form PCT/ISA/210 (second sheet) (July 1998)

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US01/31451

C. (Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	US 5,254,463 A (de BOER et al) 19 October 1993 (19.10.1993), see entire document.	1-8
---		---
Y		9-54
Y	US 5,593,972 A (WEINER et al) 14 January 1997 (14.01.1997), see entire document.	9-54
Y	US 5,482,713 A (PAOLETTI) 09 January 1996 (09.01.1996), see entire document.	9-54

Form PCT/ISA/210 (second sheet) (July 1998)

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US01/31451

Continuation of B. FIELDS SEARCHED Item 3:
STN: CAPLUS, secondary, structure, messenger#, rna#, mrna#, free energ?, thermodynamic#
WEST: secondary, structure, messenger, rna, mrna, free energ$, thermodynamic

フロントページの続き

(51)Int.Cl.[7]	FI		テーマコード（参考）
Ａ６１Ｐ 9/00	Ａ６１Ｐ 9/00		
Ａ６１Ｐ 17/00	Ａ６１Ｐ 17/00		
Ａ６１Ｐ 17/06	Ａ６１Ｐ 17/06		
Ａ６１Ｐ 19/02	Ａ６１Ｐ 19/02		
Ａ６１Ｐ 21/00	Ａ６１Ｐ 21/00		
Ａ６１Ｐ 25/00	Ａ６１Ｐ 25/00		
Ａ６１Ｐ 35/00	Ａ６１Ｐ 35/00		
Ａ６１Ｐ 37/02	Ａ６１Ｐ 37/02		
Ｃ１２Ｐ 21/02	Ｃ１２Ｐ 21/02	C	
// Ａ６１Ｋ 35/76	Ａ６１Ｋ 37/02		
Ａ６１Ｋ 38/00	Ａ６１Ｋ 37/24		
Ａ６１Ｋ 38/22	Ａ６１Ｋ 37/48		
Ａ６１Ｋ 38/43	Ａ６１Ｋ 35/76		
Ａ６１Ｋ 39/00	Ａ６１Ｋ 39/00	H	

(81)指定国 AP(GH,GM,KE,LS,MW,MZ,SD,SL,SZ,TZ,UG,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,PH,PL,PT,RO,RU,SD,SE,SG,SI,SK,SL,TJ,TM,TR,TT,TZ,UA,UG,US,UZ,VN,YU,ZA,ZW

(72)発明者 ウェイナー、デイビッド、ビー.
アメリカ合衆国、１９０６６ ペンシルバニア州、メリオン、ビーコン レーン ７１７
(72)発明者 ヤン、ジョーースン
アメリカ合衆国、１９１０４ ペンシルバニア州、フィラデルフィア、アパートメント １、サウス ４２番 ストリート １０２

Ｆターム(参考) 4B024 AA01 CA02 EA01 EA02 HA17
4B064 AG01 CA19 CC24 DA01
4C084 AA13 BA44 DB52 DC01 DC50 MA66 NA10 ZA022 ZA362 ZA662
ZA892 ZA942 ZB072 ZB152 ZB262 ZC062 ZC352
4C085 AA03 BB01 BB11 DD62 EE01 GG02 GG03 GG04 GG06
4C087 BC83 CA12 NA10 ZA02 ZA36 ZA66 ZA89 ZA94 ZB07 ZB15
ZB26 ZC06 ZC35

专利名称(译)	高表达基因		
公开(公告)号	JP2004510440A	公开(公告)日	2004-04-08
申请号	JP2002532657	申请日	2001-10-04
[标]申请(专利权)人(译)	宾夕法尼亚大学		
申请(专利权)人(译)	宾夕法尼亚大学的受托人		
[标]发明人	ウェイナーデイビッドビー ヤンジョースン		
发明人	ウェイナー、デイビッド、ビー. ヤン、ジョー-スン		
IPC分类号	C12N15/09 A61K35/76 A61K38/00 A61K38/16 A61K38/22 A61K38/43 A61K39/00 A61K39/12 A61K48/00 A61P1/04 A61P3/10 A61P5/14 A61P9/00 A61P17/00 A61P17/06 A61P19/02 A61P21/00 A61P25/00 A61P35/00 A61P37/02 A61P43/00 C07K14/16 C07K14/18 C12N15/67 C12P21/02 C12Q1/02 G01N33/15 G01N33/50 G01N33/53 G01N33/569		
CPC分类号	A61K38/162 A61K48/00 A61K2039/53 A61P1/04 A61P3/10 A61P5/14 A61P9/00 A61P17/00 A61P17/06 A61P19/02 A61P21/00 A61P25/00 A61P35/00 A61P37/02 A61P43/00 C07K14/005 C07K2319/02 C12N2740/16222 C12N2770/24122 C12N2770/24322 G01N33/56983 G01N2469/20 Y02A50/39 Y02A50/394		
FI分类号	C12N15/00.ZNA.A A61K48/00 A61P1/04 A61P3/10 A61P5/14 A61P9/00 A61P17/00 A61P17/06 A61P19/02 A61P21/00 A61P25/00 A61P35/00 A61P37/02 C12P21/02.C A61K37/02 A61K37/24 A61K37/48 A61K35/76 A61K39/00.H		
F-TERM分类号	4B024/AA01 4B024/CA02 4B024/EA01 4B024/EA02 4B024/HA17 4B064/AG01 4B064/CA19 4B064/CC24 4B064/DA01 4C084/AA13 4C084/BA44 4C084/DB52 4C084/DC01 4C084/DC50 4C084/MA66 4C084/NA10 4C084/ZA022 4C084/ZA362 4C084/ZA662 4C084/ZA892 4C084/ZA942 4C084/ZB072 4C084/ZB152 4C084/ZB262 4C084/ZC062 4C084/ZC352 4C085/AA03 4C085/BB01 4C085/BB11 4C085/DD62 4C085/EE01 4C085/GG02 4C085/GG03 4C085/GG04 4C085/GG06 4C087/BC83 4C087/CA12 4C087/NA10 4C087/ZA02 4C087/ZA36 4C087/ZA66 4C087/ZA89 4C087/ZA94 4C087/ZB07 4C087/ZB15 4C087/ZB26 4C087/ZC06 4C087/ZC35		
代理人(译)	矢口太郎 大森纯一 山口泰明		
优先权	60/237885 2000-10-04 US		
其他公开文献	JP4472250B2		
外部链接	Espacenet		

摘要(译)

种类代码：A1本发明涉及在重组表达系统中产生蛋白质的方法，该方法在表达系统中包括翻译从异源DNA序列转录的mRNA的方法。该方法包括预测从野生型异源DNA序列转录的mRNA的二级结构的步骤。与由野生型异源DNA序列转录的mRNA的二级结构相比，由该修饰的异源DNA序列转录的mRNA的二级结构具有增加的自由能。此外，该方法包括使用修饰的异源DNA序列在重组表达系统中产生蛋白质的步骤。本发明还提供了可注射的药物组合物，其由包含修饰的编码序列的核酸分子组成。本发明还提供了编码序列。

【 図 2 】

mRNA Energy Minimization in 1st 73 Bases
WNVwt
WNVh-DJY
WNVy-DJY
ENERGY = -15.1
%wild type
ENERGY = -18.6
%human codon
ENERGY = -8.3
%yeast codon