(19)日本国特許庁(JP)　(12)公表特許公報(A)　(11)特許出願公表番号
特表2005-507641
(P2005-507641A)
(43)公表日　平成17年3月24日(2005.3.24)

(51)Int.Cl.[7]	F I		テーマコード（参考）
C12N 15/09	C12N 15/00	ZNAA	4B024
C12P 21/00	C12P 21/00	C	4B064

審査請求 未請求　予備審査請求 有　（全 152 頁）

(21)出願番号	特願2002-578470(P2002-578470)
(86)(22)出願日	平成14年2月1日(2002.2.1)
(85)翻訳文提出日	平成15年8月4日(2003.8.4)
(86)国際出願番号	PCT/US2002/003055
(87)国際公開番号	WO2002/079468
(87)国際公開日	平成14年10月10日(2002.10.10)
(31)優先権主張番号	60/266,386
(32)優先日	平成13年2月2日(2001.2.2)
(33)優先権主張国	米国(US)
(31)優先権主張番号	60/268,785
(32)優先日	平成13年2月14日(2001.2.14)
(33)優先権主張国	米国(US)

(71)出願人　500273137
ラージ　スケール　バイオロジー　コーポレイション
アメリカ合衆国，カリフォルニア　９５６８８，ベイカビル　スウィート　１００，ベイカ　バレー　パークウェイ　３３３３
(74)代理人　100099759
弁理士　青木　篤
(74)代理人　100077517
弁理士　石田　敬
(74)代理人　100087413
弁理士　古賀　哲次
(74)代理人　100117019
弁理士　渡辺　陽一

最終頁に続く

(54)【発明の名称】ヘテロ二本鎖の相補性を増大させる方法

(57)【要約】
ヘテロ二本鎖ポリヌクレオチドの相補性を増大させるin vitroでの方法を記載する。当該方法は、ミスマッチを有するポリヌクレオチドの二本鎖を形成するために、逆鎖のアニーリングを使用する。ヘテロ二本鎖ポリヌクレオチドは、鎖開裂活性、３’→５’エキソヌクレアーゼ活性、及びポリメラーゼ活性を有し、相補性のパーセンテージが当該ヘテロ二本鎖内で増大するのに十分な時間を与える有効量の酵素と組み合わされる。全てのヘテロ二本鎖ポリヌクレオチドが必ずしも全てのミスマッチを相補に解消させるわけではない。生じたポリヌクレオチドは任意にライゲーションされる。複数の変異体ポリヌクレオチドが生じる。逆鎖のそれぞれが鋳型と合わされて他方の鎖を記録する部位で、ヘテロポリヌクレオチド鎖の相補性のパーセンテージの結果が増大する。親のポリヌクレオチドは、異種の鎖をアニーリングする前に、フラグメントに開裂される必要がない。従って、再会合の必要がない。

【特許請求の範囲】
【請求項１】
少なくとも２つの相補的でないヌクレオチドの塩基対を有する少なくとも１つのヘテロ二本鎖ポリヌクレオチドから、配列変異体を生成するin vitroの方法であって、
a.少なくとも１つのヘテロ二本鎖ポリヌクレオチドを調製し；
b.前記ヘテロ二本鎖ポリヌクレオチドを、有効量のCEL I、T4 DNAポリメラーゼ、及びT4 DNAリガーゼと組み合わせ；そして
c.相補性のパーセンテージが増大するのに十分な時間を与え、１又は複数の変異体が生成させること、
を含んで成る方法。
【請求項２】
少なくとも２つの相補的でないヌクレオチドの塩基対を有する少なくとも１つのヘテロ二本鎖ポリヌクレオチドから、配列変異体を生成するin vitroの方法であって、
a.少なくとも１つのヘテロ二本鎖ポリヌクレオチドを調製し；
b.前記ヘテロ二本鎖ポリヌクレオチドを、有効量のエキソヌクレアーゼ活性、ポリメラーゼ活性、及び鎖開裂活性を有する１又は複数の因子と組み合わせ；そして
c.相補性のパーセンテージが増大するのに十分な時間を与え、１又は複数の変異体を生成させること、
を含んで成る方法。
【請求項３】
前記ヘテロ二本鎖ポリヌクレオチドが環状である、請求項２に記載の方法。
【請求項４】
前記ヘテロ二本鎖ポリヌクレオチドが直鎖状である、請求項２に記載の方法。
【請求項５】
前記ヘテロ二本鎖ポリヌクレオチドがレプリコンである、請求項３又は４に記載の方法。
【請求項６】
前記変異体が異なる程度の相補性を有する、請求項２に記載の方法。
【請求項７】
エキソヌクレアーゼ活性、ポリメラーゼ活性、及び鎖開裂活性を有する前記因子が順次添加される、請求項２に記載の方法。
【請求項８】
エキソヌクレアーゼ活性、ポリメラーゼ活性、及び鎖開裂活性を有する前記因子が同時に添加される、請求項２に記載の方法。
【請求項９】
段階(b)において更にリガーゼ活性を含んで成る、請求項２に記載の方法。
【請求項１０】
(d)リガーゼを添加する、
段階を更に含んで成る、請求項２に記載の方法。
【請求項１１】
エキソヌクレアーゼ活性、ポリメラーゼ活性、リガーゼ活性、及び鎖開裂活性を有する前記因子が順次添加される、請求項２に記載の方法。
【請求項１２】
エキソヌクレアーゼ活性、ポリメラーゼ活性、リガーゼ活性、及び鎖開裂活性を有する前記因子が同時に添加される、請求項２に記載の方法。
【請求項１３】
前記リガーゼがT4 DNAリガーゼ、E.コリのDNAリガーゼ、又はTaq DNAリガーゼである、請求項９に記載の方法。
【請求項１４】
鎖開裂活性を有する前記因子が酵素である、請求項２に記載の方法。
【請求項１５】

鎖開裂活性を有する前記因子がミスマッチエンドヌクレアーゼである、請求項２に記載の方法。
【請求項１６】
鎖開裂活性を有する前記因子が、CEL I、T4エンドヌクレアーゼVII、T7エンドヌクレアーゼI、S1ヌクレアーゼ、BAL-31ヌクレアーゼ、FEN1、クリーバーゼ(cleavase)、膵臓デオキシリボヌクレアーゼI、SPヌクレアーゼ、リョクトウヌクレアーゼ、及びヌクレアーゼP1から成る群から選択される、請求項２に記載の方法。
【請求項１７】
鎖開裂活性を有する前記因子が化学物質である、請求項２に記載の方法。
【請求項１８】
鎖開裂活性を有する前記因子が、過マンガン酸カリウム、酢酸テトラエチルアンモニウム、立体的に嵩高い、光で活性化されうるＤＮＡ干渉因子、[Rh(bpy)2(chrysi)]3+、四酸化オスミウムとピペリジン、及びヒドロキシルアミンとピペリジンから成る群から選択される、請求項２に記載の方法。
【請求項１９】
鎖開裂活性を有する前記因子が電離放射線、又は動力学的放射線である、請求項２に記載の方法。
【請求項２０】
ポリメラーゼ活性を有する前記因子がT4 DNAポリメラーゼである、請求項２に記載の方法。
【請求項２１】
ポリメラーゼ活性を有する前記因子がT7 DNAポリメラーゼである、請求項２に記載の方法。
【請求項２２】
ポリメラーゼ活性及び３’→５’エキソヌクレアーゼ活性の両方を有する前記因子が、T4 DNAポリメラーゼ、T7 DNAポリメラーゼ、E.コリのPol 1、又はPfu DNAポリメラーゼである、請求項２に記載の方法。
【請求項２３】
ポリメラーゼ活性及び５’→３’エキソヌクレアーゼ活性の両方を有する前記因子がE.コリのPol 1である、請求項２に記載の方法。
【請求項２４】
前記有効量の鎖開裂活性、及びエキソヌクレアーゼ活性／ポリメラーゼ活性及びリガーゼ活性が、CEL I、T4 DNAポリメラーゼ、及びT4 DNAリガーゼによって提供される、請求項２に記載の方法。
【請求項２５】
前記有効量の鎖開裂活性、及びエキソヌクレアーゼ活性／ポリメラーゼ活性及びリガーゼ活性が、CEL I、T7 DNAポリメラーゼ、及びT4 DNAリガーゼによって提供される、請求項２に記載の方法。
【請求項２６】
前記有効量の鎖開裂活性、及びエキソヌクレアーゼ活性／ポリメラーゼ活性及びリガーゼ活性が、T4エンドヌクレアーゼVII、T4 DNAポリメラーゼ、及びT4 DNAリガーゼによって提供される、請求項２に記載の方法。
【請求項２７】
配列群の多様性を増大させるin vitroの方法であって、少なくとも１つのヘテロ二本鎖ポリヌクレオチドを調製し；前記ヘテロ二本鎖ポリヌクレオチドを、有効量の３’→５’エキソヌクレアーゼ活性、ポリメラーゼ活性、及び鎖開裂活性を有する１又は複数の因子と組み合わせ；そして、相補性のパーセンテージが増大するのに十分な時間を与え、前記群の多様性を増大させること、
を含んで成る方法。
【請求項２８】

前記ヘテロ二本鎖ポリヌクレオチドが環状である、請求項２７に記載の方法。
【請求項２９】
前記ヘテロ二本鎖ポリヌクレオチドが直鎖状である、請求項２７に記載の方法。
【請求項３０】
前記ヘテロ二本鎖ポリヌクレオチドがレプリコンである、請求項３又は４に記載の方法。
【請求項３１】
前記変異体が異なる程度の相補性を有する、請求項２７に記載の方法。
【請求項３２】
３’→５’エキソヌクレアーゼ活性、ポリメラーゼ活性、及び鎖開裂活性を有する前記因子が順次添加される、請求項２７に記載の方法。
【請求項３３】
３’→５’エキソヌクレアーゼ活性、ポリメラーゼ活性、及び鎖開裂活性を有する前記因子が同時に添加される、請求項２７に記載の方法。
【請求項３４】
リガーゼを添加することを更に含んで成る、請求項２７に記載の方法。
【請求項３５】
前記リガーゼがT4 DNAリガーゼ、E.コリのDNAリガーゼ、又はTaq DNAリガーゼである、請求項９に記載の方法。
【請求項３６】
鎖開裂活性を有する前記因子が酵素である、請求項２７に記載の方法。
【請求項３７】
鎖開裂活性を有する前記因子がミスマッチエンドヌクレアーゼである、請求項２７に記載の方法。
【請求項３８】
鎖開裂活性を有する前記因子が、CEL I、T4エンドヌクレアーゼVII、T7エンドヌクレアーゼI、S1ヌクレアーゼ、BAL-31ヌクレアーゼ、FEN1、クリーバーゼ(cleavase)、膵臓デオキシリボヌクレアーゼI、SPヌクレアーゼ、リョクトウヌクレアーゼ、及びヌクレアーゼP1から成る群から選択される、請求項２７に記載の方法。
【請求項３９】
鎖開裂活性を有する前記因子が化学物質である、請求項２７に記載の方法。
【請求項４０】
鎖開裂活性を有する前記因子が、過マンガン酸カリウム、酢酸テトラエチルアンモニウム、立体的に嵩高い、光で活性化されうるＤＮＡ干渉因子、[Rh(bpy)2(chrysi)]3+、四酸化オスミウムとピペリジン、及びヒドロキシルアミンとピペリジンから成る群から選択される、請求項２７に記載の方法。
【請求項４１】
鎖開裂活性を有する前記因子が電離放射線、又は動力学的放射線である、請求項２７に記載の方法。
【請求項４２】
ポリメラーゼ活性を有する前記因子がT4 DNAポリメラーゼである、請求項２７に記載の方法。
【請求項４３】
ポリメラーゼ活性を有する前記因子がT7 DNAポリメラーゼである、請求項２７に記載の方法。
【請求項４４】
ポリメラーゼ活性及び３’→５’エキソヌクレアーゼ活性の両方を有する前記因子が、T4 DNAポリメラーゼ、T7 DNAポリメラーゼ、E.コリのPol 1、又はPfu DNAポリメラーゼである、請求項２７に記載の方法。
【請求項４５】
ポリメラーゼ活性及び５’→３’エキソヌクレアーゼ活性の両方を有する前記因子がE.コ

リのPol 1である、請求項２７に記載の方法。
【請求項４６】
前記有効量の鎖開裂活性、及びエキソヌクレアーゼ活性／ポリメラーゼ活性及びリガーゼ活性が、CEL I、T4 DNAポリメラーゼ、及びT4 DNAリガーゼによってそれぞれ提供される、請求項２７に記載の方法。
【請求項４７】
前記有効量の鎖開裂活性、及びエキソヌクレアーゼ活性／ポリメラーゼ活性及びリガーゼ活性が、CEL I、T7 DNAポリメラーゼ、及びT4 DNAリガーゼによってそれぞれ提供される、請求項２７に記載の方法。
【請求項４８】
前記有効量の鎖開裂活性、及びエキソヌクレアーゼ活性／ポリメラーゼ活性及びリガーゼ活性が、CEL I、T7 DNAポリメラーゼ、及びT4 DNAリガーゼによってそれぞれ提供される、請求項２７に記載の方法。
【請求項４９】
前記有効量の鎖開裂活性、及びエキソヌクレアーゼ活性／ポリメラーゼ活性及びリガーゼ活性が、T4エンドヌクレアーゼVII、T4 DNAポリメラーゼ、及びT4 DNAリガーゼによってそれぞれ提供される、請求項２７に記載の方法。
【請求項５０】
配列群の多様性を増大させるin vitroの方法であって、
a.少なくとも１つのヘテロ二本鎖ポリヌクレオチドを調製し；
b.前記ヘテロ二本鎖ポリヌクレオチドを、有効量のCEL I、T4 DNAポリメラーゼ、及びT4 DNAリガーゼと組み合わせ；そして
c.相補性のパーセンテージが増大するのに十分な時間を与え、当該群の多様性を増大させること、
を含んで成る方法。
【請求項５１】
所望の機能的特性をコードするポリヌクレオチドを得る方法であって、
a.少なくとも１つのヘテロ二本鎖ポリヌクレオチドを調製し；
b.前記ヘテロ二本鎖ポリヌクレオチドを、有効量のエキソヌクレアーゼ活性、ポリメラーゼ活性、及び鎖開裂活性を有する１又は複数の因子と組み合わせ；
c.ヘテロ二本鎖ポリヌクレオチド鎖間の相補性のパーセンテージが増大するのに十分な時間を与え、当該群の多様性を増大させ；そして
d.所望の機能的特性について、変異体群をスクリーニングし、又は選択すること、
を含んで成る方法。
【請求項５２】
所望の機能的特性をコードするポリヌクレオチドを得る方法であって、
a.少なくとも１つのヘテロ二本鎖ポリヌクレオチドを調製し；
b.前記ヘテロ二本鎖ポリヌクレオチドを、有効量のエキソヌクレアーゼ活性、ポリメラーゼ活性、及び鎖開裂活性を有する１又は複数の因子と組み合わせ；
c.ヘテロ二本鎖ポリヌクレオチド鎖間の相補性のパーセンテージが増大するのに十分な時間を与え、当該群の多様性を増大させ、
d.DNAをRNAに変換し；そして
e.所望の機能的特性について、変異体群をスクリーニングし、又は選択すること、
を含んで成る方法。
【請求項５３】
所望の機能的特性を有するポリペプチドを得る方法であって、
a.少なくとも１つのヘテロ二本鎖ポリヌクレオチドを調製し；
b.前記ヘテロ二本鎖ポリヌクレオチドを、有効量のエキソヌクレアーゼ活性、ポリメラーゼ活性、及び鎖開裂活性を有する１又は複数の因子と組み合わせ；
c.ヘテロ二本鎖ポリヌクレオチド鎖間の相補性のパーセンテージが増大するのに十分な時

間を与え、
d.前記ヘテロ二本鎖ポリヌクレオチドをRNAに、そして前記RNAをポリペプチドに変換し；そして
e.所望の機能的特性について、ポリペプチド変異体群をスクリーニングし、又は選択すること、
を含んで成る方法。
【請求項５４】
所望の機能的特性コードするポリヌクレオチドを得る方法であって、
a.少なくとも１つのヘテロ二本鎖ポリヌクレオチドを調製し；
b.前記ヘテロ二本鎖ポリヌクレオチドを、有効量のエキソヌクレアーゼ活性、ポリメラーゼ活性、及び鎖開裂活性を有する１又は複数の因子と組み合わせ；
c.ヘテロ二本鎖ポリヌクレオチド鎖間の相補性のパーセンテージが増大するのに十分な時間を与え、
d.所望の機能的特性を有する変異体群をスクリーニングし、又は選択し；
e.一本鎖ポリヌクレオチドを得るために前記変異体群を変性させ；
f.少なくとも１つの第二のヘテロ二本鎖ポリヌクレオチドを形成するために、前記一本鎖ポリヌクレオチドをアニーリングし；
g.前記第二ヘテロ二本鎖ポリヌクレオチドを、有効量のエキソヌクレアーゼ活性、ポリメラーゼ活性、及び鎖開裂活性を有する１又は複数の因子と組み合わせ；そして
h.ヘテロ二本鎖ポリヌクレオチド鎖間の相補性のパーセンテージが増大するのに十分な時間を与えること、
を含んで成る方法。
【請求項５５】
前記ヘテロ二本鎖ポリヌクレオチドが95%超同一である、請求項５４に記載の方法。
【請求項５６】
前記ヘテロ二本鎖ポリヌクレオチドが約95%同一である、請求項５４に記載の方法。
【請求項５７】
前記ヘテロ二本鎖ポリヌクレオチドが約90%同一である、請求項５４に記載の方法。
【請求項５８】
前記ヘテロ二本鎖ポリヌクレオチドが約85%同一である、請求項５４に記載の方法。
【請求項５９】
前記ヘテロ二本鎖ポリヌクレオチドが約80%同一である、請求項５４に記載の方法。
【請求項６０】
前記ヘテロ二本鎖ポリヌクレオチドが約75%同一である、請求項５４に記載の方法。
【請求項６１】
前記ヘテロ二本鎖ポリヌクレオチドが約1000kbである、請求項５４に記載の方法。
【請求項６２】
前記ヘテロ二本鎖ポリヌクレオチドが約10,000kbである、請求項５４に記載の方法。
【請求項６３】
前記ヘテロ二本鎖ポリヌクレオチドが約100,000kbである、請求項５４に記載の方法。
【請求項６４】
配列群の多様性を増大させるために使用されるキットであって、少なくとも１つのヘテロ二本鎖ポリヌクレオチドを調製し；前記ヘテロ二本鎖ポリヌクレオチドを、有効量の３’→５’エキソヌクレアーゼ活性、ポリメラーゼ活性、及び鎖開裂活性を有する１又は複数の因子と組み合わせ；そして、相補性のパーセンテージが増大するのに十分な時間を与え、前記群の多様性を増大させること、
を含んで成るキット。
【請求項６５】
リガーゼ活性を有することを更に含んで成る、請求項６４に記載のキット。
【発明の詳細な説明】

【技術分野】
【０００１】
本発明は、広く分子生物学に、そして更に具体的には関連する核酸分子群を生成する方法に関する。
【背景技術】
【０００２】
DNAシャッフリングは、２ないしそれ以上のDNA配列間での組み換え体を得て、それらを加速度的に進化させるための強力なツールである。DNAシャッフリング法のための親DNA又はインプットDNAは、典型的には、野生型以上の、複数の向上した特徴を有する所定の遺伝子の突然変異体又は変異体である。DNAシャッフリング産物は、新規配列の組み合わせから生じる追加的な作用又は相乗作用について解析されうる、親DNA由来の遺伝子配列の本質的にランダムな再編成のプールを表す。
【０００３】
反復的な配列の再編成は、適当な特性を有する変異体のみが、自身の遺伝物質を次の世代の産生に寄与することを許される、進化の過程に似ている。最適化された変異体は、DNAシャッフリングを媒介した配列の再編成、それに続く性能の漸進的な向上の試験を介して産生する。再編成及び試験の追加のサイクルは、前記過程の前の回で同定された遺伝的な向上の新規組み合わせを含む遺伝子の産生を導く。利益のある遺伝的な変化を再編成させ、そして組み合わせることは、最適化された配列が、個別に生成し、そしてあらゆる可能性のある配列の組み合わせをスクリーニングする必要性なしに、最適化した配列を生じさせる。
【０００４】
これは、既に改良された配列に対するその後の改良が大部分セレンディピティからもたらされるランダム突然変異誘発とはっきりと異なる。例えば、増強された特性の所望の集合を有するタンパク質を得るために、種々の有益な変異の組み合わせを含む突然変異体を同定する必要があることがある。これらの有益な遺伝的変化を組み合わせるための方法が利用可能でない場合、更なるランダム突然変異誘発が必要とされる。しかしながら、ランダム突然変異誘発は、多数の突然変異体を産生させ、そしてスクリーニングする繰り返しのサイクルを必要とし、結果として退屈で且つ非常に労働力を要する方法である。更に、配列が不所望な作用を有する突然変異を被る割合は、配列の情報量によって増大する。故に、情報量、ライブラリーのサイズ、及び突然変異の割合が増大するにつれ、有益なミューテーションに対する有害なミューテーションの割合は増大し、更なる改良の選択を漸次的に覆い隠す。最後に、複数のコンピューターシミュレーションは、点変異誘発だけでは、継続的で且つ劇的な配列の進化に必要とされる大規模な阻止的変化を可能にするのに、しばしばあまりにも緩やかすぎるであろうことを示唆している。
【０００５】
ランダム突然変異誘発に使用される、多数の異なる技術が存在している。例えば、ある方法は、ライブラリーフォーマットにおいて突然変異体遺伝子を作るために、エラープローンポリメラーゼ連鎖反応(PCR)を利用する(Cadwell and Joyce, 1992; Gram et al., 1992).。別の方法は、最適化されるべき特異的領域が合成的に突然変異したオリゴヌクレオチドで置換される、カセット式変異誘発(Arkin and Youvan, 1992; Delagrave et al., 1993; Delagrave and Youvan, 1993; Goldman and Youvan, 1992; Hermes et al., 1990 ; Oliphant et al., 1986; Stemmer et al., 1993)である。
【０００６】
エラープローンPCRは、配列全体に低レベルの点変異をランダムに誘導するために、低フィデリティーの重合条件を使用する。しかしながら、この方法に対する制限は、公開されたエラープローンPCRのプロトコールが、ポリメラーゼの低い前進性に苦しみ、このことが、このアプローチを、平均サイズの遺伝子においてランダム突然変異誘発を生むのに効率的でなくする。
【０００７】

オリゴヌクレオチド指定ランダム突然変異誘発において、短い配列は、合成的に変異したオリゴヌクレオチドで置換される。離れたミューテーションの組み合わせを生み出すために、異なる部位が異なるオリゴヌクレオチドによって同時に扱われなければならない。全ての部位を満たすのに必要なライブラリーサイズと比較して、限定されたこの方法で得られるライブラリーサイズは、多数の選択が最適化に必要とされることを意味する。合成オリゴヌクレオチドによる突然変異誘発は、各選択段階後の個々のクローンの配列決定、それに続いてそれらをファミリーに分類し、単一のファミリーを任意に選択し、そしてそれを共通モチーフへと縮小することを必要とする。そのようなモチーフは、再合成されて、単一の遺伝子に再挿入され、更に選択される。この段階は統計学的なボトルネックとなるものであり、労働力を要し、そして多数の突然変異誘発にとって実践的ではない。

【０００８】

これらの理由のために、エラープローンPCR及びオリゴヌクレオチド指定突然変異誘発は、比較的少ない配列の変化のサイクルを必要とする突然変異プロトコール、例えば配列の微調整のために使用されうるが、多数の突然変異及び選択のサイクルを必要とする、特に大きい遺伝子配列に対する手順にとってのそれらの有用性において制限される。

【０００９】

上述のように、ランダムに変異した遺伝子から改良された遺伝子産物を産生するためのこれまでの方法は、利用が限定されている。広範なランダムに再編成した遺伝子配列を産生するための、１つの認識されている方法は、長いヌクレオチドをより短い断片へと開裂する酵素を使用する。続いて、開裂因子は、遺伝物質から分離され、そして当該物質はポリヌクレオチド鎖として再編成できるように増幅され、ここで、それらの再編成はランダム又は特異的な順番のものである((Stemmer, 1994a ; Stemmer, 1994b), 米国特許第5,605,793号, 米国特許第5,811,238号, 米国特許第5,830,721号, 米国特許第5,928,905号, 米国特許第6,096,548号, 米国特許第6,117,679号, 米国特許第6,165,793号, 米国特許第6,153,410号)。この変法は、再編成前にフラグメントを生成するために、プライマー及び限定されたポリメラーゼ伸長を使用する(米国特許第5,965,408号, 米国特許第6,159,687号)。

【００１０】

しかしながら、両方の方法とも限界を有する。これらの方法は、技術的に複雑であるという欠点がある。このことは、これらの方法の利用可能性を、熟練のスタッフを抱える施設に限定させる。更に、フラグメントからの分子の再編成に起因する複雑さ、例えば意図していない突然変異誘発及びサイズが増大する巨大な標的分子の再編成の困難性の増大が存在しており、これらのことは、長いポリヌクレオチド鎖を再編成させる場合、これらの方法の有用性を制限する。

【００１１】

断片化及び再編成を基にした遺伝子シャッフリングの、これらの方法の別の制限は、親の鋳型ポリヌクレオチドが次第に異質となる場合に遭遇する。それらの方法のアニーリング段階において、小さいポリヌクレオチドフラグメントは、適切にアニーリングするために、塩基対合の相互作用から生じる安定化する力に依存する。アニーリングの小さい領域が、それらの短い長さに起因して、限定された安定化する力を有するので、高度に相補的な配列のアニーリングは、更に多岐にわたる配列よりも恵まれている。そのような場合、これらの方法は、特定の親の鋳型に由来する相補的な一本鎖のアニーリングに起因して、親の鋳型ポリヌクレオチドを再生する傾向が強い。それ故に、親の鋳型は、本質的にそれら自身を再会合してライブラリーにおいて未変化のポリヌクレオチドのバックグラウンドを作り出し、このことが、組み換え分子の検出をより困難にする。この問題は、親の鋳型が更に異質になるにつれ、すなわち、親の鋳型間の配列の同一性のパーセンテージが低下するにつれ、ますます深刻となる。この結果はKikuchi等によって証明され(Gene 243 : 133-137,2000)、彼らは、Patten et al., 1997; Crameri et al., 1998; Harayama, 1998; Kumamaru et al., 1998; Chang et al., 1999 ; Hansson et al., 1999によって報告されたファミリーシャッフリング法を用いて、xylEとnahHとの間での組み換え体を産生するこ

とを試みた。Kikuchi等は、本質的に全く組み換え体が産生されなかったこと明らかにした(＜1%)。彼らはまた、一本鎖DNAの断片化及び再会合によってキメラ遺伝子の形成を向上させる方法を開示している。この方法を用いて、彼らは、１４％の割合のキメラ遺伝子を手に入れたが、残りの８４％が親配列であった。

【００１２】

組み換え体の低効率の回収特性は、低いパーセンテージの配列同一性を有する親の鋳型、すなわちより多様な親の鋳型から新規ポリヌクレオチドを産生する場合に、これらの方法の有用性を制限する。従って、これらの必要性に対処する遺伝子配列を産生する方法についての必要性が存在する。

【００１３】

本発明は前述の必要性を満たす方法を提供し、そして更には関連する利益も提供する。

【００１４】

本発明の要約

本発明は、ヘテロ二本鎖分子を形成し、そして次にミスマッチの部位の配列情報が一方の鎖から他方に移されるように当該ミスマッチを取り扱うことによって、関連するポリヌクレオチド間の突然変異を再編成させる方法を提供する。１つの好ましい態様において、当該ミスマッチは、ミスマッチニッキング酵素、dNTPの存在下で３’から５’にプルーフリーディングする活性を有するポリメラーゼ、及びリガーゼを含む反応液中でヘテロ二本鎖分子をインキュベートすることによって取り扱われる。これらの各活性は、所定のミスマッチ部位において、ヘテロ二本鎖がニックされ、不対塩基が切り出され、鋳型として反対の鎖を用いて置き換えられ、そしてニックが塞がれるように一斉に働く。アウトプットのポリヌクレオチドはクローニング前に増幅されるか、又は直接クローニングされ、そして向上した特性について試験される。ミスマッチの解消の再編成及び試験の追加のサイクルは更なる向上をもたらす。

【００１５】

定義

本明細書で使用する場合、用語「増幅」は、ポリヌクレオチドのコピー数が増大する過程を言及する。

【００１６】

本明細書で使用する場合、「アニーリング」は、少なくとも部分的に相補的なヌクレオチド配列のハイブリダイゼーションによる少なくとも部分的に二本鎖の核酸の形成を言及する。部分的に二本鎖の核酸は、より大きな核酸鎖に対するより小さな核酸鎖のハイブリダイゼーション（ここで、より小さな核酸は、より大きな核酸の一部と１００％同一である）に起因することがある。部分的に二本鎖の核酸はまた、１００％の同一性を有さないが、特定のハイブリダイゼーション条件下でハイブリダイズするのに十分な相同性を有する２つの核酸鎖のハイブリダイゼーションに起因することがある。

【００１７】

本明細書で使用する場合、「クランプ」は、例えばPCRプライマー内へのクランプ配列の組み込みによって、ポリヌクレオチドの一方の末端に付加される独特のヌクレオチド配列を言及する。クランプ配列は、異なる親由来の鎖のハイブリダイゼーションから生じるポリヌクレオチド（すなわちヘテロ二本鎖分子）の増幅のみを可能にし、それによって、既に説明されているような完全長のハイブリッド産物の産生を保証することを目的とする（Skarfstad, J. Bact, vol 182, No 11, P. 3008-3016)。

【００１８】

本明細書で使用する場合、用語「開裂」は、酵素によってポリヌクレオチドを消化すること、又はポリヌクレオチド内のホスホジエステル結合を他の方法で破壊することを意味する。

【００１９】

本明細書で使用する場合、用語「相補的な塩基対」は、一方の鎖にあるアデニンが他方の鎖にあるチミン（又はウラシル）と対になり、そして一方の鎖にあるシトシンが他方の鎖

にあるグアニンと対になるような、二本鎖中のDNA（又はRNA）の対応物を言及する。

【００２０】

本明細書で使用する場合、用語「～と相補的」は、本明細書では、相補的な配列が引用ポリヌクレオチド配列の全部又は一部の反転した相補鎖と同一であること、又は一方の鎖にある各ヌクレオチドが、反対の鎖にあるヌクレオチド、若しくはその類似体と塩基対を形成することができることを意味する。例示のために、ヌクレオチド配列「TATAC」は、引用配列「GTATA」と相補的である。

【００２１】

本明細書で使用する、「変性する」又は「変性した」は、核酸に関して使用する場合、二本鎖核酸の、一本鎖核酸への変換を言及する。二本鎖核酸を変性させる方法は当業者に周知であり、そして、例えば塩基対合を不安定にする因子を添加し、温度を増大し、塩を減らし、又はそれらを組み合わせることを含む。これらの要因は、鎖の相補性に従い、すなわち、鎖が１００％相補的であるか、あるいは１又は複数の相補的でないヌクレオチドを有するか、に従い適用される。

【００２２】

本明細書で使用する場合、用語「所望の機能的特性」は、限定されないが、ポリペプチドをコードすること、連結したポリヌクレオチドの転写を促進すること；タンパク質に結合すること、ウイルスベクターの機能を向上させること、等であって、選択可能であり、又はそれらについてスクリーニング可能であるものを含む、表現型の特性を意味する。そのような所望な機能的特性を有するポリヌクレオチドは、限定しないが、適当な植物、動物、菌、酵母又は細菌の発現ベクターからの発現、トランスジェニック植物、動物又は微生物を形成するための組み込み、リボザイムの発現等を含む、多くの方法において使用されうる。

【００２３】

本明細書で使用する場合、用語「DNAシャッフリング」は、本明細書では実質的に相同であるが同一でない配列間での組み換えを示唆するために使用した。

【００２４】

本明細書で使用する場合、用語「有効量」は、所望の活性を提供する因子に必要とされる当該因子の量を言及する。本発明の場合、この決定は当業者に周知である。

【００２５】

本明細書で使用する場合、用語「エキソヌクレアーゼ」は、ポリヌクレオチド鎖の末端から１つずつヌクレオチドを開裂する酵素、すなわち、ポリヌクレオチド分子の３’又は５’末端からホスホジエステル結合を加水分解する酵素を言及する。そのようなエキソヌクレアーゼは、限定しないが、T4 DNAポリメラーゼ，T7 DNA ポリメラーゼ，E. coli Pol 1，及びPfu DNA ポリメラーゼを含む。用語「エキソヌクレアーゼ」活性は、エキソヌクレアーゼに関連する活性を言及する。３’から５’方向に加水分解するエキソヌクレアーゼは、３’→５’のエキソヌクレアーゼ活性を有すると言われる。同様に、５’→３’活性を揺するエキソヌクレアーゼは、５’→３’エキソヌクレアーゼ活性を有すると言われる。いくつかのエキソヌクレアーゼ、例えばE. コリのPol I.が、３’→５’、５’→３’の両方の活性を有することに注意しなければならない。

【００２６】

本明細書で使用する場合、「ミスマッチ解消による遺伝子再編成(GRAMMR)」は、ヘテロ二本鎖分子を形成し、そして情報が一方の鎖から他方に移されるようにミスマッチを取り扱うことによって関連するポリヌクレオチド鎖間の配列のばらつきを再編成させるための方法を言及する。

【００２７】

本明細書で使用する場合、「グラニュラリティー」は、隣接配列として、鋳型のポリヌクレオチド鎖から第二のポリヌクレオチド鎖に移される核酸の配列情報量を言及する。本明細書で使用する場合、「鋳型配列」は、GRAMMRによる処理が、鋳型鎖から第二鎖への遺伝情報の移動をもたらすような、第二のポリヌクレオチド配列と部分的に相補的な、第一の

一本鎖ポリヌクレオチド配列を言及する。

【００２８】

鋳型鎖から移動した配列情報の単位が大きいほど、グラニュラリティーは高い。鋳型鎖から移動した配列情報のブロックが小さいほど、グラニュラリティーはより低いかより細かい。より低いグラニュラリティーは、DNAシャッフリング又は再編成の方法が、鋳型鎖から第二鎖へと、遺伝情報のより小さな不連続のブロックを移動させることができることを示唆している。より低いグラニュラリティーを有するDNAのシャッフリング又は再編成の方法の利点は、より小さな核酸配列を他のものから分離し、又は配列情報を移動させることができることである。主に高いグラニュラリティーを戻すDNAのシャッフリング又は再編成の方法は、より小さな核酸配列を他のものから容易に分離することはできない。

【００２９】

本明細書で使用する場合、用語「ヘテロ二本鎖ポリヌクレオチド」は、一本鎖、典型的には別々の鎖（ここで、当該鎖は同一ではない）をアニーリングすることによって形成される二重らせんポリヌクレオチドを言及する。ヘテロ二本鎖ポリヌクレオチドは、一本鎖のループ又はバブルとして存在する不対の領域を有することがある。ヘテロ二本鎖ポリヌクレオチド領域はまた、１つの一本鎖ポリヌクレオチドによって形成されることがあり、ここで、部分的な自己相補性はステムループ構造の形成を可能にし、ここで、当該鎖のアニーリング部分は同一ではない。

【００３０】

本明細書で使用する場合、用語「ヘテロ二本鎖DNA」は、一本鎖、典型的には別々の鎖（ここで、当該鎖は同一ではない）をアニーリングすることによって形成されるDNAの二重らせんを言及する。ヘテロ二本鎖DNAは、一本鎖のループ又はバブルとして存在する不対の領域を有することがある。ヘテロ二本鎖DNAはまた、１つの一本鎖ポリヌクレオチドによって形成されることがあり、ここで、部分的な自己相補性はステムループ構造の形成を可能にし、ここで、当該鎖のアニーリング部分は同一ではない。

【００３１】

本明細書で使用する場合、用語「相同」は、１つの一本鎖の核酸配列が、少なくとも部分的に相補的な一本鎖の核酸配列とハイブリダイズしうることを意味する。ハイブリダイゼーションの程度は、配列間の同一性の程度及びハイブリダイゼーション条件、例えば後述するような温度及び塩濃度を含む多数の要因に依存することがある。

【００３２】

核酸は、それらが共通の祖先配列から天然又は人工的に由来する場合、「相同」である。自然な進化の間に、これが起こるのは、２又はそれ以上の祖先配列が長時間かけて親配列から分岐する場合であり、すなわちこれは突然変異及び天然の選択に起因する。人工的な条件のもと、分岐は、例えば２つの基本的な経路の一方で生じる。最初に、所定の配列は、例えば典型的なクローニングの間に、子孫核酸配列を生み出すように起こることで、別の配列と人工的に組み換えられることがあり、又は処置の配列は化学的に修飾されることがあり、又は生じた分子を修飾するために他の方法により操作されることがある。あるいは、核酸は、選択された親核酸配列から配列が変化している核酸を合成することによって新規合成されることがある。２つの核酸の祖先について、明示的な知識がない場合、相同性は、２つの配列間の配列の比較によって推測される。２つの核酸配列が、各核酸の重大な部分に及ぶ配列類似性を示す場合、２つの核酸は共通の祖先を有することが推定される。相同性を確立する配列類似性の正確なレベルは、様々な要因に依存して当業界で変化する。

【００３３】

本発明の開示のために、２つの核酸は、それらが、GRAMMR媒介型の情報転送を２つの核酸分子間で起こさせるのに十分な配列同一性を共有する場合に相同であるとみなされる。

【００３４】

本明細書で使用する場合、用語「同一」又は「同一性」は、２つの核酸配列が同一配列又は相補配列を有することを意味する。このように、「同一性の領域」は、ポリヌクレオチ

ドの範囲又は領域あるいはポリヌクレオチド全体が別のポリヌクレオチド領域に対して同一又は相補的であることを意味する。

【００３５】

本明細書で使用する場合、用語「相補的なパーセンテージの増大」は、ヘテロ二本鎖分子内の相補的な塩基対のパーセンテージがより大きくなることを意味する。

【００３６】

本明細書で使用する場合、「リガーゼ」は、核酸中の破壊されたホスホジエステル結合を再連結する酵素を言及する。

【００３７】

本明細書で使用する場合、用語「ミスマッチ」は、正常な塩基対合の相互作用を形成することができない塩基対を言及する（すなわち、「A」と「T(又はU)」、又は「G」と「C」以外のもの）。

【００３８】

本明細書で使用する場合、用語「ミスマッチ解消」は、ミスマッチの塩基対の、相補的な塩基対への変換を言及する。

【００３９】

本明細書で使用する場合、用語「突然変異」は、野生型又は参照核酸配列の配列の変化、あるいはポリペプチドの配列の変化を意味する。そのような突然変異は、点突然変異、例えばトランジション又はトランスバージョンであってもよい。突然変異は、欠失、挿入又は重複であってもよい。

【００４０】

本明細書で使用する場合、用語「ニックトランスレーション」は、５’→３’のエキソヌクレアーゼ活性と５’→３’ポリメラーゼ活性の組み合わせが、二本鎖ポリヌクレオチドの一本鎖の破壊の位置（「ニック」）を５’→３’方向に動かすポリメラーゼの特性を言及する。

【００４１】

本明細書で使用する場合、用語「核酸」又は「核酸分子」は、ポリヌクレオチド、例えばデオキシリボ核酸（ＤＮＡ）又はリボ核酸（ＲＮＡ）を意味し、そして一本鎖核酸及び二本鎖の核酸及びオリゴヌクレオチドを包含する。本発明において有用な核酸は、ゲノムＤＮＡ、ｃＤＮＡ、ｍＲＮＡ及び合成オリゴヌクレオチドを含み、そしてセンス鎖、アンチセンス鎖、又はその両方を表すことがある。核酸は通常、４つの天然のヌクレオチド、アデニン、グアニン、シトシン、及びチミジン／ウリジンが組み込まれている。本発明の核酸はまた、他の天然、非天然のヌクレオチドが組み込まれていることがあり、ヌクレオチド誘導体が、ポリメラーゼによって、所望のポリヌクレオチド産物を生成するのに十分な効率でポリヌクレオチド内に組み込まれうる限り、それらの誘導体を含む。

【００４２】

本明細書で使用する場合、「親核酸」は、出発時の部分的に相補な核酸群と１００％同一な配列を有する二本鎖核酸を言及する。親核酸は、例えば図２の例示において、部分的に相補な核酸の組み合わせ１＋／４－又は２－／３＋が本発明の方法において出発時の集団として使用された場合、核酸Ｘ及びＹを含む。

【００４３】

本明細書で使用する場合、「部分的に相補」は、別の核酸に対して実質的に相補的な配列を有するが、少なくとも２又はそれ以上のヌクレオチドが他の核酸と異なる二本鎖核酸を言及する。本明細書で使用する場合、「部分的に相補的な核酸群」は、実質的に相補的な配列を有する核酸を含んで成るが、当該群の任意な他のメンバーにとって正確な相補配列を有する核酸のない群を言及する。本明細書で使用する場合、部分的に相補的な核酸群の任意なメンバーは、２又はそれ以上のヌクレオチドが当該群の別の核酸、又はそれらの相補鎖と異なる。従って、部分的に相補的な核酸は、厳密に相補的な配列、すなわち、１００％の相補性を有する相補配列を含む群を特異的に除外する。それ故に、そのような部分的に相補的な核酸群の各メンバーは、両方の鎖を含む２又はそれ以上のヌクレオチドが、

当該群の他のメンバーと異なる。一方の鎖は、上側の鎖(top strand)と表され、そしてその相補鎖は下側の鎖(bottom strand)と表される。本明細書で使用する場合、「上側」の鎖は５’から３’方向に読まれるポリヌクレオチドを言及し、そして「下側」のものはその相補鎖を言及する。配列は上側又は下側の鎖として言及されるが、そのような表記が、溶液中で上側又は下側の鎖として鎖を固定する配向がないので相補鎖を識別することを意図していることは理解されよう。

【００４４】

例えば、２つの核酸のメンバーを含む群は２つの二本鎖の核酸に由来することがあり、これは、一本鎖の部分的に相補的な核酸群を生成するために４本の鎖のいずれかを用いる可能性を有する。本発明の部分的に相補的な核酸群を得るために使用されうる２つの核酸鎖の組み合わせの例を図２に示す。部分的に相補的な核酸群の中の可能性のあるメンバーである２つの核酸配列は、「Ｘ」（AGATCAATTG; 配列番号１）及び「Ｙ」（AGACCGATTG; 配列番号２）と表される（図２Ａ）。当該核酸配列は、２つの位置が異なる（「★」で示した４位及び６位）。核酸Ｘ及びＹの「上側」の鎖は、それぞれ「１＋」及び「３＋」と表され、そして核酸Ｘ及びＹの「下側」の鎖は、それぞれ「２－」及び「４－」と表される。

【００４５】

図２Ｂは、４本の核酸鎖の可能性のある組み合わせを示す。６つの可能性のある鎖の組み合わせのうち、１＋／２－、１＋／４－、２－／３＋、又は３＋／４－の組み合わせのみが、部分的に相補的な核酸群の中の必要な上側及び下側の鎖を含んで成る。これらの上側／下側の配列の組み合わせのうち、１＋／４－又は２－／３＋のみが、２つの異なる分子の部分的に相補的な核酸群の例を含んで成るのは、これらの組み合わせだけが、少なくとも１つのヌクレオチドが異なる相補配列を有するためである。残りの組み合わせ１＋／２－及び２＋／４－は、厳密に相補的な配列を含み、そしてそれ故に本発明の部分的に相補的な核酸群を含まない。

【００４６】

２つの異なる分子の群の上述の例において、核酸分子の部分的に相補的な核酸分子群は、１又は複数のヌクレオチドが異なるが、同一のセンス、例えば１＋／３－又は２－／４－である鎖の組み合わせを除外した。しかしながら、そのような同一の核酸鎖の組み合わせは、当該群が少なくとも１本の下側の鎖及び少なくとも１つの上側の鎖を含む限り、より大きな群に含まれうると理解される。例えば、５＋及び６－の鎖を有する第三の核酸「Ｚ」が含まれる場合、１＋／３＋／６－又は２－／４－／５＋の組み合わせは、部分的に相補的な核酸群を含んで成る。同様に、核酸の任意なメンバー並びにそれらの相当する上側及び下側の鎖は、部分的に相補的な核酸群を生成するために、当該群が少なくとも１つの上側の鎖及び少なくとも１つの下側の鎖を含む限り、且つ当該群が厳密に相補的なメンバーを含まない限り組み合わせられうる。

【００４７】

本発明の核酸群は、約３又はそれ以上、約４又はそれ以上、約５又はそれ以上、約６又はそれ以上、約７又はそれ以上、約８又はそれ以上、約９又はそれ以上、約１０又はそれ以上、約１２又はそれ以上、約１５又はそれ以上、約２０又はそれ以上、約２５又はそれ以上、約３０又はそれ以上、約４０又はそれ以上、約５０又はそれ以上、約７５又はそれ以上、約１００又はそれ以上、約１５０又はそれ以上、約２００又はそれ以上、約２５０又はそれ以上、約３００又はそれ以上、約３５０又はそれ以上、約４００又はそれ以上、約４５０又はそれ以上、約５００又はそれ以上、あるいは約１０００又はそれ以上の異なる核酸分子であってもよい。群はまた、約２０００又はそれ以上、約５０００又はそれ以上、約1×10^4又はそれ以上、約1×10^5又はそれ以上、約1×10^6又はそれ以上、約1×10^7又はそれ以上、あるいは約$x \times 10^8$又はそれ以上、約５又はそれ以上の異なる核酸分子を含んでもよい。当業者は、本発明の方法に含めるために、所望の再編成の実験結果及び本明細書に開示されているような利用可なスクリーニング方法に依存して、所望の群を容易に決定することができる。

【００４８】
本明細書で使用する場合、「ポリメラーゼ」は、ヌクレオチドのポリマー、すなわちポリヌクレオチドの形成を触媒する酵素を言及する。本発明において有用なポリメラーゼは任意の生物又は起源に由来することがあり、動物、植物、細菌又はウイルスのポリメラーゼを含む。ポリメラーゼは、ＤＮＡポリメラーゼ、ＲＮＡポリメラーゼ、ＲＮＡをＤＮＡに転写することができる逆転写酵素であってもよい。
【００４９】
本明細書で使用する場合、「プルーフリーディング」は、ヌクレオチド、例えばミスマッチのヌクレオチドが３’→５’エキソヌクレアーゼ活性によって除去され、そして、典型的に塩基対合したヌクレオチドによって置換されうる酵素の特性を説明する。
【００５０】
本明細書で使用する場合、「組み換え」ポリヌクレオチドは、少なくとも２つの異なるポリヌクレオチド由来の配列情報を含んで成るポリヌクレオチドを言及する。
【００５１】
本明細書で使用する場合、用語「関連ポリヌクレオチド」は、当該ポリヌクレオチドの範囲又は領域が同一であり、且つ当該ポリヌクレオチドの範囲及び領域が同一でないことを意味する。
【００５２】
本明細書で使用する場合、用語DNAの「再編成」は、実質的に相同であるが同一でない配列間の配列のばらつきの再分布を示唆するために本明細書で使用される。
【００５３】
本明細書で使用する場合、用語「レプリコン」は、一定の長さのポリヌクレオチドを含む複製の遺伝的単位及びその複製開始部位を言及する。
【００５４】
本明細書で使用する場合、用語「配列の多様性」は、同一でないポリヌクレオチドの存在量を言及する。用語「群における配列の多様性の増大」は、群における同一でないポリヌクレオチドの存在量が増大することを意味する。
【００５５】
本明細書で使用する場合、用語「配列変異体」は、参照分子と比較して、１又は複数の配列の差違を有する分子（ＤＮＡ、ＲＮＡポリペプチド等）を言及する。例えば、GRAMMRの過程の間にヘテロ二本鎖分子の全体で起こる別個の独立したミスマッチの解消の事象の合計が、その分子全体の配列情報の再編成をもたらす。配列情報は、「配列変異体」の複雑なライブラリーを生成するために、様々な組み合わせで再編成する。
【００５６】
本明細書で使用する場合、「鎖の開裂活性」又は「開裂」は、ニックの形成におけるような、ポリヌクレオチドの主鎖中のホスホジエステル結合の破壊を言及する。鎖の開裂活性は酵素的な因子(そのような因子は、限定しないが、CEL I, T4エンドヌクレアーゼVII, T7エンドヌクレアーゼ I, S1ヌクレアーゼ, BAL-31ヌクレアーゼ, FEN1, クリーバーゼ(cleavase)、膵臓デオキシリボヌクレアーゼI、SPヌクレアーゼ、リョクトウヌクレアーゼ、及びヌクレアーゼP1を含む）；化学物質（そのような化学物質は、限定しないが、過マンガン酸カリウム、酢酸テトラエチルアンモニウム、立体的に嵩高い、光で活性化されうるＤＮＡ干渉因子、[Rh(bpy)2(chrysi)]3+、四酸化オスミウムとピペリジン、及びヒドロキシルアミンとピペリジン、を含む）；又は電離放射線、又は動力学的放射線(kinetic radiation)の形態のエネルギーによって提供されうる。
【００５７】
本明細書で使用する場合、用語「十分な時間」は、所望の産物を与えるために反応又は過程に必要とされる時間を言及する。本発明の場合、十分な時間の決定は当業者の知識の範囲内にある。「十分な時間」が、実施者の望みに応じて、反応の機能性、又は所望の産物の失に影響を与えることなく広範に変化しうることに注意すること。
【００５８】

本明細書で使用する場合、用語「野生型」が、核酸フラグメントが突然変異を全く含まないことを意味する。「野生型」タンパク質は、当該タンパク質が天然で見られる活性レベルで活性であり、そして典型的には天然で見られるアミノ酸配列であることを意味する。ある観点において、用語「野生型」又は「親配列」は、本発明の操作の前の出発時の配列又は参照配列を示唆することがある。

【００５９】

本明細書で使用されるポリペプチドの表記において、標準的な使用法及び慣習に従い、左手の方向はアミノ末端の方向であり、そして右手の方向はカルボキシル末端の方向である。同様に、特に断らない限り、一本鎖のポリヌクレオチド配列の左手の末端は５’末端であり；二本鎖のポリヌクレオチド配列の左手の方向は５’方向として言及される。新生ＲＮＡ転写物の５’から３’への付加の方向は、転写方向と言及される。

【００６０】

本発明の詳細な説明

本発明は、少なくとも２つの相補的でないヌクレオチドの塩基対を有する少なくとも１つのヘテロ二本鎖ポリヌクレオチドから配列変異体を生成するin vitroの方法であって；少なくとも１つのヘテロ二本鎖ポリヌクレオチドを調製し；前記ヘテロ二本鎖ポリヌクレオチドを、有効量のエキソヌクレアーゼ活性、ポリメラーゼ活性及び鎖の開裂活性を有する１又は複数の因子と組み合わせ；そして相補性のパーセンテージを増大させるのに十分な時間を与えることを含んで成り、少なくとも１つ又は複数の変異体が生成する、方法を提供する。

【００６１】

本発明の別の観点は、ヘテロ二本鎖ポリヌクレオチドが環状、直鎖又はレプリコンである場合のものである。

【００６２】

本発明の別の観点は、所望な変異体が異なる程度の相補性を有する場合のものである。

【００６３】

本発明の別の観点は、エキソヌクレアーゼ活性、ポリメラーゼ活性、及び鎖開裂活性が連続的に又は同時に加えられる場合のものである。

【００６４】

本発明の別の観点は、T4 DNAリガーゼ、E.コリDNAリガーゼ、又はTaq DNAリガーゼのような因子によって提供され、リガーゼ活性の追加を提供する。

【００６５】

本発明の別の観点は、鎖の開裂活性が、酵素、例えばCEL I, T4エンドヌクレアーゼVII, T7エンドヌクレアーゼ I, S1ヌクレアーゼ, BAL-31ヌクレアーゼ, FEN1, クリーバーゼ(cleavase)、膵臓デオキシリボヌクレアーゼI、SPヌクレアーゼ、リョクトウヌクレアーゼ、及びヌクレアーゼP1；化学物質、例えば過マンガン酸カリウム、酢酸テトラエチルアンモニウム、立体的に嵩高い、光で活性化されうるＤＮＡ干渉因子、[Rh(bpy)2(chrysi)]3+、四酸化オスミウムとピペリジン、及びヒドロキシルアミンとピペリジン；又はエネルギー形態、例えば電離放射線、又は動力学的放射線によって提供されるものである。

【００６６】

本発明の別の観点は、ポリメラーゼ活性がPolベータによって提供されるものである。

【００６７】

本発明の別の観点は、ポリメラーゼ活性及び３’→５’エキソヌクレアーゼ活性の両方がT4 DNAポリメラーゼ、T7 DNAポリメラーゼ、E.コリ Pol 1, 又はPfu DNAポリメラーゼによって提供されるものである。

【００６８】

本発明の別の観点は、ポリメラーゼ活性及び５’→３’エキソヌクレアーゼ活性の両方を有する因子がE.コリ Pol 1であるものである。

【００６９】

本発明の別の態様は、有効量の鎖の開裂活性、並びにエキソヌクレアーゼ活性／ポリメラ

ーゼ活性及びリガーゼ活性が、CEL I、T4 DNAポリメラーゼ、及びT4 DNAリガーゼによって提供されるものである。

【００７０】

本発明の別の観点は、有効量の鎖の開裂活性、並びにエキソヌクレアーゼ活性／ポリメラーゼ活性及びリガーゼ活性が、CEL I、T7 DNAポリメラーゼ、及びT4 DNAリガーゼによって提供されるものである。

【００７１】

本発明の別の態様は、配列群の多様性を増大させるin vitroの方法であって、少なくとも１つのヘテロ二本鎖ポリヌクレオチドを調製し；前記ヘテロ二本鎖ポリヌクレオチドを、有効量のエキソヌクレアーゼ活性、ポリメラーゼ活性及び鎖の開裂活性を有する１又は複数の因子と組み合わせ；そして相補性のパーセンテージを増大させるのに十分な時間を与えて当該群の多様性が増大すること、を含んで成る方法を提供する。

【００７２】

本発明の別の態様は、所望の機能的特性をコードするポリヌクレオチドを得る方法であって、少なくとも１つのヘテロ二本鎖ポリヌクレオチドを調製し；前記ヘテロ二本鎖ポリヌクレオチドを、エキソヌクレアーゼ活性、ポリメラーゼ活性及び鎖の開裂活性を有する、有効量の１又は複数の因子と組み合わせ；ヘテロ二本鎖のポリヌクレオチド鎖間の相補性のパーセンテージを増大させるのに十分な時間を与えて当該群の多様性が増大し；そして所望の機能的特性について変異体群スクリーニングし又は選択すること、を含んで成る方法を提供する。

【００７３】

本発明の別の態様は、所望の機能的特性をコードするポリヌクレオチドを得る方法であって、少なくとも１つのヘテロ二本鎖ポリヌクレオチドを調製し；前記ヘテロ二本鎖ポリヌクレオチドを、エキソヌクレアーゼ活性、ポリメラーゼ活性及び鎖の開裂活性を有する、有効量の１又は複数の因子と組み合わせ；ヘテロ二本鎖のポリヌクレオチド鎖間の相補性のパーセンテージを増大させるのに十分な時間を与えて当該群の多様性が増大し；ＤＮＡをＲＮＡに変換し；そして所望の機能的特性についてリボ核酸の変異体群をスクリーニングし又は選択すること、を含んで成る方法を提供する。

【００７４】

本発明の更に別の態様は、所望の機能的特性を有するポリペプチドを得る方法であって、少なくとも１つのヘテロ二本鎖ポリヌクレオチドを調製し；前記ヘテロ二本鎖ポリヌクレオチドを、エキソヌクレアーゼ活性、ポリメラーゼ活性及び鎖の開裂活性を有する、有効量の１又は複数の因子と組み合わせ；ヘテロ二本鎖のポリヌクレオチド鎖間の相補性のパーセンテージを増大させるのに十分な時間を与え；前記ヘテロ二本鎖ポリヌクレオチドをＲＮＡに、そして前記ＲＮＡをポリペプチドに変換し；そして前記所望の機能的特性についてポリペプチド変異体群をスクリーニングし又は選択すること、を含んで成る方法を提供する。

【００７５】

本発明の更に別の態様は、所望の機能的特性をコードするポリヌクレオチドを得る方法であって、任意に９５％、９０％、８５％、８０％、又は７５％同一であり、且つ約１００００ＫＢ、１０，０００ＫＢ、又は１００，０００ＫＢのサイズである、少なくとも１つのヘテロ二本鎖ポリヌクレオチドを調製し；前記ヘテロ二本鎖ポリヌクレオチドを、エキソヌクレアーゼ活性、ポリメラーゼ活性及び鎖の開裂活性を有する、有効量の１又は複数の因子と組み合わせ；ヘテロ二本鎖ポリヌクレオチド鎖間の相補性のパーセンテージを増大させるのに十分な時間を与え；前記ヘテロ二本鎖ポリヌクレオチドをＲＮＡに、そして前記ＲＮＡをポリペプチドに変換し；前記所望の機能的特性についてポリペプチド変異体群をスクリーニングし又は選択し；一本鎖ポリヌクレオチドを得るために、前記変異体群を変性させ；少なくとも１つの第二のヘテロ二本鎖ポリヌクレオチドを形成するために、前記一本鎖ポリヌクレオチドをアニーリングし；前記第二ヘテロ二本鎖ポリヌクレオチドを、エキソヌクレアーゼ活性、ポリメラーゼ活性及び鎖の開裂活性を有する、有効量の１又

は複数の因子と組み合わせ；そしてヘテロ二本鎖ポリヌクレオチド鎖間の相補性のパーセンテージを増大させるのに十分な時間を与えること、を含んで成る方法を提供する。

【００７６】

本発明は、典型的に増幅され、そして／あるいはクローニングされたポリヌクレオチドの形態にある、向上したポリヌクレオチド配列又は向上したポリヌクレオチド配列群を産生する方法であって、それによって向上したポリヌクレオチド配列が、選択され、又はスクリーニングされうる少なくとも１つの所望の表現型の特徴（例えば、ポリペプチドをコードし、連結したポリヌクレオチドの転写を促進し、タンパク質に結合し、ウイルスベクターの機能を向上させる等）を有する、方法に関する。そのような所望のポリヌクレオチドは、多くの方法で、例えば適当な植物、動物、真菌、酵母、又は細菌からの発現、トランスジェニック植物、動物又は微生物を形成するための組み込み、リボザイムの発現等で使用されうる。

【００７７】

GRAMMRは、ヘテロ二本鎖ＤＮＡの鎖が、アニーリング、その後のin vitro反応でのミスマッチの解消によって作り出される方法を提供する。この反応は、ミスマッチ部位の又はミスマッチ付近の一方の鎖又は他方の鎖の開裂で開始し、これに、ミスマッチ塩基のその鎖からの除去、そして他方の鎖の配列に対して鋳型となるヌクレオチドで、生じたギャップを埋めるためのポリマー化、が続く。生じたニックは、主鎖を再連結するためのライゲーションによって塞がれうる。ヘテロ二本鎖分子全体で起こる別々の独立したミスマッチを解消する事象の集合は、その分子全体の配列情報の再編成をもたらす。当該配列情報は、配列変異体の複合ライブラリーを生成するために、様々な組み合わせで再編成する。

【００７８】

GRAMMRの１つの態様において、突然変異体のライブラリーは、当業界で知られている任意な方法、例えば。変異原性ＰＣＲ、化学的突然変異誘発等、その後の所望の特性を有する突然変異体についてのスクリーニング又は選択によって生成する。ＤＮＡは、選択された突然変異体から調製される。突然変異体のＤＮＡは、混合され、一本鎖に変性され、そしてアニーリングさせられる。ハイブリダイズする部分的に相補的な鎖は、ミスマッチ部位において、塩基対合していないヌクレオチドを有する。CEL I(Oleykowski et al., 1998; Yang et al., 2000)又は類似のミスマッチに関する活性による処理は、各ミスマッチの、一方又は他方のポリヌクレオチド鎖の３’側にニッキングをもたらす。（尚、CEL Iは、挿入／欠失の再編成をもたらす挿入／欠失の３’側にニックをいれることができる。）３’→５’のエキソヌクレアーゼ（「プルーフリーディング」）活性を含むポリメラーゼ（例えば、T4 DNA Pol）の存在は、ミスマッチの除去を可能にし、そしてそれに続き、５’→３’のポリメラーゼ活性は、鋳型として他の鎖を用いてギャップを埋める。５’→３’のエキソヌクレアーゼ活性及び鎖の置換活性を欠くポリメラーゼはギャップを埋め、そして最初のCEL I開裂部位に位置するＤＮＡの５’末端に達したときポリマー化を終了し、その結果、配列のごく短い部分を再合成する。あるいは、合成された部分の長さは、５’－３’エキソヌクレアーゼ活性を含むポリメラーゼで当該反応を抑えることによって調整され；このニックトランスレーション活性は、より長い領域を行き来することができ、それにより、鋳型の鎖から移動したより長い情報の部分をもたらす。続いて、ＤＮＡリガーゼ（例えば、T4 DNAリガーゼ）は、修復された鎖のリン酸の主鎖を元に戻すことによってニックを塞ぐことができる。この工程は、所定のヘテロ二本鎖ＤＮＡ分子の多くの部位及びいずれかの鎖で、同時に起こりうる。この結果は、出発配列群よりも多様性のある配列変異体群を与える、インプットの鎖の間の配列の差異のランダム化による。これらアウトプットのポリヌクレオチドは、直接的に適当なベクター内にクローニングされ、又はそれらはクローニング前にＰＣＲによって増幅されうる。あるいは、当該反応は、二本鎖環状プラスミド分子内のヘテロ二本鎖領域又はGRAMMR反応後に適当な宿主内に直接導入されうる他の適当な領域上で実施されうる。別の選択肢において、アウトプットのポリヌクレオチドはＲＮＡのポリヌクレオチドに転写され、そして、例えばウイルスベクター転写プラスミドの場合において、例えば植物上への植物ウイルスベクターの接種によって直接的

に使用される。生じたクローンは、所望の特性の向上について選択又はスクリーニングにかけられる。続いて、全体の工程が、更なる向上のために、選択されたクローンを用いて１又は複数回繰り返されることがある。

【００７９】

アウトプットのポリヌクレオチドが直接クローニングされる場合、クローニング宿主の複製時に、同一細胞内に２つの異なるプラスミドをもたらしうることを主張する、不完全に解消した分子の可能性が存在する。これらのプラスミドは、潜在的に混合型のプラスミドを生じさせることがある。そのような可能性を避けることが望ましい場合、アウトプットのポリヌクレオチド分子は、複製／解消をさせるために宿主内で生育され、当該ポリヌクレオチドは単離され、そして新規宿主細胞を再形質転換させられうる。

【００８０】

別の態様において、一分子あたり２つ以上の親に由来する配列のインプットが望ましい場合、上記手順は、アウトプットのポリヌクレオチドの任意のクローニングの前に繰り返し実施される。GRAMMR処理の後、二本鎖ポリヌクレオチドは変性し、アニーリングさせられ、そしてミスマッチの解決の工程が繰り返される。所望の回数のそのようなサイクルの後、アウトプットのポリヌクレオチドが直接的にクローニングされ、適当なベクター内に導入され、あるいはそれらはクローニング前のＰＣＲによって増幅されうる。生じたクローンは、所望の特性の向上についての選択又はスクリーニングにかけられる。

【００８１】

別の態様において、「分子の戻し交配(molecular backcross)」は、所望の突然変異からの突然変異のバックグラウンドを排除するのを助けるために実施される。所望の突然変異体のＤＮＡのプールは、当該方法を実施するために、適当な比率の野生型ＤＮＡと混合されうる。クローンは向上について選択され、プールされ、そして更なる有意な変化がなくなるまで再び野生型に戻されうる。

【００８２】

当該工程の効率は、ヘテロ二本鎖分子に関して出発時の群をエンリッチメントにする様々な方法によって向上し、それにより、未変化の親の型のアウトプット分子の数が減少する。ミスマッチのハイブリッドは、アプタマー、色素、又はミスマッチのＤＮＡに結合する他の因子を用いてアフィニティー精製されうる。好ましい態様は、Mutsタンパク質アフィニティーマトリックス(Wagner et al., Nucleic Acids Res. 23 (19): 3944-3948 (1995); Su et al., Proc. Natl. Acad. Sci. (U. S. A.), 83: 5057-5061 (1986))又はT4 エンドヌクレアーゼ VIIの、ミスマッチに結合するが開裂しない突然変異体(Golz and Kemper,Nucleic Acids Research, 1999; 27: e7)の使用である。

【００８３】

１つの態様において、当該手順は、インプットのポリヌクレオチドが一本鎖の各配列変異体の一本鎖から成るように修飾される。例えば、逆鎖の一本鎖ＤＮＡは、部分的に相補的な一本鎖分子を精製するために、非対称ＰＣＲによって、異なる親の配列から産生される。当該鎖とヘテロ二本鎖を作製するための互いの鎖とのアニーリングは、例１に記載のように実施される。あるいは、一本鎖のＤＮＡは、ラムダエキソヌクレアーゼを用いる、各親の二本鎖ＤＮＡの一方の鎖を優先的に消化し、続いて残っている鎖を互いにアニーリングさせることによって産生しうる。この態様において、アニーリングする鎖は、１００％の相補的な鎖を有さず、これは再アニーリングすることを示す。従って、アウトプットのポリヌクレオチド間には、未修飾のポリヌクレオチド、すなわち「親ポリヌクレオチド」のより少ないバックグラウンドが存在しており、これがより高い効率で再編成する配列変異体へと導く。この増大した効率は、選択よりもむしろスクリーニングが所望のポリヌクレオチドの試験に適用される状況において特に価値がある。

【００８４】

ヘテロ二本鎖を形成する別の方法は、二本鎖の親のＤＮＡを混合し、この鎖を解離するために変性させ、そしてヘテロ二本鎖と親のホモ二本鎖の群を精製するために、一本鎖ＤＮＡを互いにアニーリングさせることである。当該ヘテロ二本鎖は、続いて、ヘテロ二本鎖

捕捉法、例えばMutS又は開裂しないT4エンドヌクレアーゼVII突然変異体を用いる上述の方法によって選択的にエンリッチメントにすることができる。あるいは、当該群の親のヘテロ二本鎖分子は、ヘテロ二本鎖は開裂しないが、親のホモ二本鎖分子は開裂するような、ミスマッチの部位を覆う制限酵素によって開裂されうる。開裂されないヘテロ二本鎖ＤＮＡは、続いて、例６に記載のように完全長プラスミドのヘテロ二本鎖ＤＮＡ上で間全長プラスミドを精製するために実施されたように、アガロースゲル中でのサイズ分画によって単離されうる。続いて、それらの完全長ヘテロ二本鎖プラスミド分子の環化は、ＤＮＡリガーゼとのインキュベーションによってもたらされた。

【００８５】

別の態様において、親、又はインプットの、二本鎖ポリヌクレオチドは、「クランプ」配列の付加によって修飾される。一方のインプットポリヌクレオチド又はポリヌクレオチドのプールは、５’プライマーの独特の配列の付加によるＰＣＲによって増幅される。他方のインプットポリヌクレオチド又はプールは、３’プライマーの独特な配列の付加によるＰＣＲによって増幅される。クランプ配列は、注目の遺伝子の５’末端のための独特な制限酵素部位及び３’末端のための別のものを含むように設計され、その結果、GRAMMR再編成産物をクローニングする段階で、第一ポリヌクレオチド（又はプール）由来の５’クランプ及び第二ポリヌクレオチド（又はプール）を有する産物のみが、クローニングに適切な末端を有する。あるいは、GRAMMR再編成産物は、同様の結果を達成するために、５’及び３’クランプの独特な配列を用いてＰＣＲ増幅されうる。従って、未修飾ポリヌクレオチド、すなわち、アウトプットのポリヌクレオチドクローンの間の「親のポリヌクレオチド」のバックグラウンドがより少なくなり、より高効率の再編成する配列変異体をもたらす。この増大した効率は、選択よりもむしろスクリーニングが所望のポリヌクレオチドの試験に適用される状況において特に価値がある。任意に、クランププライマー配列と相補的なオリゴヌクレオチドプライマーが、GRAMMR反応に添加され、その結果、いずれかの親が上側の鎖として働くことがあり、それにより、相互のヘテロ二本鎖がともにミスマッチの解決反応に参加することを可能にする。

【００８６】

環状ヘテロ二本鎖ポリヌクレオチドを産生するための別の方法は、親の二本鎖ＤＮＡが上述のような末端のクランプ配列を有し、ヘテロ二本鎖の一方の末端から伸びている一本鎖のクランプ配列が、ヘテロ二本鎖の他方の末端から伸びている一本鎖のクランプ配列と相補的であるように実施される。これらの相補的な一本鎖のクランプはアニーリングさせられ、それによってヘテロ二本鎖ＤＮＡ分子を環化する。同一配列の再アニーリングから生じる親のホモ二本鎖は、たった１つのクランプ配列を有し、それ故に、環化が起こりうるそれらの末端において相補的な一本鎖配列を有さない。更に、ＤＮＡポリメラーゼ及びＤＮＡリガーゼは、それぞれ、環状分子内の任意なギャップを埋めるために、そして主鎖のニックを塞ぐために使用されることがあり、その結果、共有結合で閉じた環状ヘテロ二本鎖分子群をもたらす。更なる変性条件にかけられた場合でも、共有結合で閉じた環状ヘテロ二本鎖分子がそれらの構成成分である鎖に解離しないので、変性、環化及びライゲーションの工程が、より多くの直線状の親の二本鎖が閉環状のヘテロ二本鎖に変換するために繰り返されうる。

【００８７】

別の態様において、一本鎖の環状ファージミドＤＮＡの領域は、関連しているが同一でない直鎖ＤＮＡとハイブリダイズすることができ、これは、続いてポリメラーゼ、例えばT7 DNAポリメラーゼ又はT4 DNAポリメラーゼ＋T4の遺伝子32のタンパク質を用いて伸長され、続いてＤＮＡ間で異なる部位にあるヘテロ二本鎖領域を有する環状の二本鎖分子を得るために生じたニックにおいてライゲーションされうる。GRAMMRは、配列が再編成した分子のライブラリーを得るために、この分子上で実施されうる。

【００８８】

あるいは、プラスミドの主鎖と比較して極性の、逆鎖の２つの一本鎖環状ファージミドＤＮＡ、及び再編成の標的である親遺伝子配列は互いにアニーリングされる。広範なミスマ

ッチの領域が、ファージf1起源の配列が存在する場所で生じる。しかしながら、GRAMMR処理時に、この広範なミスマッチの領域は、機能f1の起源を回復する親の型の配列のいずれかに戻ることがある。これらの２回変化させられた分子はまた、注目の親の遺伝子をコードする鎖間の異なる部位にミスマッチの領域を含む。続いて、GRAMMRは、配列が再編成された分子のライブラリーを得るために、この分子上で実施されうる。

【００８９】

前節で論じたように、出発時のＤＮＡ又はインプットＤＮＡは、任意な数の形態のものであってもよい。例えば、インプットＤＮＡは、例１で教示するように、完全長、一本鎖、そして反対のセンスのものでもよい。あるいは、インプットＤＮＡはまた、完全長の鎖のフラグメントであってもよい。インプットＤＮＡは、一方が又は両方ともが二本鎖であっても、あるいは、例えばメチル化、ホスホロチオエート結合、ペプチド－核酸、一方又は両方の鎖におけるＲＮＡの置換等によって修飾されてもよい。二本鎖の一方の鎖は、両方の鎖に沿って連続的であるか、連続的ではないが近接しているか、重複部分で断続的となっているか、又はギャップで断続的となっていてもよい。

【００９０】

GRAMMRはまた、ＤＮＡの断片化及び再編成を基にしたＤＮＡシャッフリングのスキームにかけられてもよい。例えば、遺伝子フラグメントが、遺伝子再編成の過程において、変性、アニーリング、及び伸長にかけられる方法において、GRAMMRは中間段階として利用されうる。

【００９１】

１つのそのような態様において、遺伝子、又は変異体の遺伝子由来のＤＮＡは、酵素的、機械的又は化学的手段によって断片化され、そして任意に前記フラグメントのサイズ範囲は、アガロースゲル上での分離のような手段によって単離される。出発時のポリヌクレオチド、例えば野生型のもの、所望の変異体、あるいはそれらのプールは、前記フラグメントに添加され、そして混合物は変性させられ、そしてアニーリングさせられる。アニーリングしたポリヌクレオチドは、鋳型として無傷の鎖を用いて、一本鎖のギャップを埋めるために、ポリメラーゼで処理される。生じた部分的に相補的な二本鎖は、ミスマッチ部位で塩基対合していないヌクレオチドを有する。CEL I(Oleykowski et al., 1998; Yang et al., 2000)による処理は、各ミスマッチの、一方又は他方のポリヌクレオチドの３’側にニッキングをもたらす。プルーフリーディング活性を提供する３’→５’のエキソヌクレアーゼを含むポリメラーゼ、例えばDNA Pol I、T4 DNA Pol 1は、ミスマッチの除去を可能にし、続いて、５’→３’のポリメラーゼ活性は、鋳型として他の鎖を用いてギャップを埋める。ＤＮＡリガーゼ、例えばT4 DNAリガーゼは、修復された鎖のリン酸基を修復することによって、ニックを塞ぐことができる。この結果は、潜在的に向上した特性を有するアウトプット鎖を精製するために、インプット鎖間の配列の変化のランダム化によるものである。これらのアウトプットポリヌクレオチドは、適当なベクター内に直接クローニングされることがあり、あるいはそれらはクローニング前にＰＣＲによって増幅されることがある。生じたクローンは、所望の特性の向上について選択又はスクリーニングにかけられる。

【００９２】

１つのそのような態様において、遺伝子又は突然変異体の遺伝子プール由来のＤＮＡは、酵素的、機械的又は化学的手段によってフラグメント化され、あるいはフラグメントは親の鋳型にアニーリングしたランダムなオリゴヌクレオチドの限定的な伸長によって生成し（米国特許第5,965,408号）、そして任意に、前記フラグメントのサイズ範囲は、アガロースゲル上での分離のような手段によって生成する。混合物は変性させられ、そしてアニーリングさせられる。アニーリングしたポリヌクレオチドは、一本鎖のギャップを埋めるために、ポリメラーゼで任意に処理される。生じた部分的に相補的な二本鎖は、ミスマッチ部位で塩基対合していないヌクレオチドを有する。CEL I(Oleykowski et al., 1998; Yang et al., 2000)による処理は、各ミスマッチの、一方又は他方のポリヌクレオチドの３’側にニッキングをもたらす。３’→５’のエキソヌクレアーゼ（「プルーフリーディ

ング」）活性を含むポリメラーゼ、例えば、T4 DNAポリメラーゼは、ミスマッチの除去を可能にし、そしてそれに続き、５’→３’のポリメラーゼ活性は、鋳型として他の鎖を用いてギャップを埋める。任意に、ＤＮＡリガーゼ、例えばT4 DNAリガーゼは、修復された鎖のリン酸基を修復することによって、ニックを塞ぐことができる。この結果は、潜在的に向上した特性を有するアウトプット鎖を精製するために、インプット鎖間の配列の変化のランダム化によるものである。その後の変性、アニーリング、及びGRAMMRの段階は、遺伝子の再構築を可能にする。これらのＰＣＲのアウトプットは、適当なベクター内に直接クローニングされることがある。生じたクローンは、所望の特性の向上について選択又はスクリーニングにかけられる。

【００９３】

本発明の別の態様は、別の１又は複数の遺伝子フラグメントがアニーリングする、連続的なスカフォールド鎖からの出発を提供する。フラップ及びギャップは、Coco, et al., Nature Biotech 19 (01) 354; 米国特許第6,319,713号に記載のようにトリミングされ、そして埋められ、そしてGRAMMRが実施される。この方法において、GRAMMRは、鋳型鎖と、フラップ及びギャップのトリミング及びライゲーションから生じる鎖との間の配列情報の移動を可能にすることによって、更なる配列の再編成をもたらす。この方法は、特異的な配列部分を一方の連続的な鎖に組み込むこと、その後の、スカフォールドとミスマッチを形成する残基のGRAMMR、という利益を提供する。同じ配列又は遺伝子に対し同時に多くのフラグメントをアニーリングすることによって、多くの別々の部位が同時に処理され、それによって瞬時に多数の配列又は遺伝子の再編成が可能となる。Coco等によって開示された方法と異なり、本発明の態様においては、スカフォールドは分解せず、それどころか、二本鎖は直接クローニングされることがあり、あるいはクローニングの前にＰＣＲによって増幅されることがある。網羅的なミスマッチの解消は、完全に二本鎖のＤＮＡをもたらす。部分的なミスマッチの解消は、二本鎖あたり、本質的に２つの異なる再編成産物をもたらす。

【００９４】

本発明の開示から理解されるように、GRAMMRはまた、工程の中の１つの段階として関連ＤＮＡのアニーリングを含む様々な方法に適用されうる。例えば、多くの部位指定突然変異誘発は、突然変異をコードするＤＮＡ分子を、一本鎖型の環状ＤＮＡ、ファージミド又は変性プラスミドにアニーリングさせることを必要とする。これらのＤＮＡは、続いてポリメラーゼで伸長され、その後リガーゼによる処理でニックを塞がれ、更なる操作で親配列を除去され、親の遺伝的バックグラウンドに組み込まれた所望の１又は複数の突然変異が残る。これらのプロトコールは通常特異的な突然変異を特定のＤＮＡ配列に組み込むために使用されるが、GRAMMR法が２つの鎖間の配列の変化を再編成するそのような方法において産生するヘテロ二本鎖分子に適用され、それによって、再編成された遺伝的な変化を有する多様な子孫の組み合わせをもたらす。

【００９５】

別の態様は、特定の領域に対する再編成の連続的な段階を提供する。例えば、ＤＮＡフラグメントは、環状の一本鎖ファージミドＤＮＡとアニーリングされ、そしてGRAMMRが実施される。フラグメントは、それらが物理的にアウトプット材料に組み込まれるのを防ぐように処理されうる。例えば、それらは、ジデオキシ残基でそれらを伸長させないようにすることによって、３’末端側で終結させられることがある。再編成の多数の繰り返しが実施されることがあるが、最初のインプットの一本鎖ＤＮＡのクローンからわずかに修飾された分子が回収される。結果として、この再編成に使用されるＤＮＡフラグメントは、最終産物に配列情報のみを与え、そして最終的な回収可能産物内に物理的に組み込まれない。

【００９６】

約１～３個以上の部分である、重大なミスマッチの部位のみを解消させることが望ましい場合において、S1ヌクレアーゼが使用されうる。S1ヌクレアーゼは、一本鎖核酸に特異的なエンドヌクレアーゼである。それは、ＤＮＡ：ＤＮＡ又はＤＮＡ：ＲＮＡの二本鎖のミ

スマッチの塩基対の限定された領域を認識し、そして開裂することができる。少なくとも約４個の連続的な塩基対のミスマッチが、S1ヌクレアーゼによる認識及び開裂に通常必要とされる。ミスマッチの解消は、両方の鎖が開裂している場合には起こらないので、ＤＮＡは、最初のニックの後、そしてカウンターニックの前に修復されなければならない。他のヌクレアーゼは、配列、配列の前後関係、又はミスマッチのサイズに従い、開裂の特異性を特異的に調整するのに好ましいことがある。

【００９７】

尚、ミスマッチの残基を取り扱う他の手段、例えばミスマッチの化学的開裂も使用されることがある。あるいは、デオキシリボヌクレアーゼＩのよって示されるような活性又は二本鎖の領域においてのみ開裂する因子によるランダムニッキングに、ヘテロ二本鎖ＤＮＡの鎖をかけることを選択することができる。ニックの形成が２つの遺伝子間で同一の領域においてのみ起こる場合、当該反応に存在するＤＮＡリガーゼは、配列情報の純移転無しにニックを塞ぐ。しかしながら、ニックの形成がミスマッチ部位付近で起こる場合、ミスマッチの塩基は３’－５’エキソヌクレアーゼによって除去されてもよく、そしてギャップはポリメラーゼによって埋められ、続いてリガーゼによってニックを塞がれてもよい。あるいは、不均一な領域全体のニックトランスレーションの適用は、配列の再編成をもたらすことがある。これらの方法は、ＤＮＡのミスマッチの状況によって独占的に方向づけられないが、配列情報を修復された配列に移動させる役割を果たし、そしてそれにより再編成された配列をもたらす。

【００９８】

GRAMMRは、選択されたライブラリーメンバー間での組み換えを得るために、タンパク質、ペプチド、又はアプタマーディスプレイ法に使用されうる。インプットＤＮＡのフラグメント化はGRAMMRに必要ではないので、配列の最小の伸長間で配列情報を再編成することが可能と思われる。例えば、小さいペプチドをコードするＤＮＡ又は、特定の特性、例えば標的への結合について選択されたＲＮＡアプタマーが再編成されうる。選択された分子間で起こるアニーリングのために、ある程度の配列の相同性が分子間で共有されるべきであり、例えば、類似の結合活性を原因として又はコドンのゆらぎ塩基の同一性を特定のデフォルト値に偏らせることで類似性を持つランダム化された配列のセグメントの領域の、コード配列の５’と３’の領域においでである。

【００９９】

GRAMMRが実施される反応温度の操作も有用なことがある。例えば、温度が低いほどヘテロ二本鎖が安定化するのを助け、GRAMMRをより高度にミスマッチした基質で実施可能にする。同様に、鎖間の塩基対合に影響を及ぼす添加物、例えば塩、PEG、ホルムアミド等は、GRAMMRにおけるヘテロ二本鎖の安定性を変化させるのに使用されることがあり、それによって反応の結果に影響を及ぶ。

【０１００】

別の態様において、ミスマッチの二本鎖ポリヌクレオチドが生成し、アプリン酸部位又はアピリミジン酸部位（すなわち「ＡＰ部位」）を形成するためにＤＮＡグリコシラーゼで、ホスホジエステル結合を開裂するためにＡＰエンドヌクレアーゼ活性で、デオキシリブロース－リン酸分子を除去するためにホスホジエステラーゼで、単一のヌクレオチドをギャップがあるＤＮＡ鎖の３’末端に付加するためにＤＮＡポリメラーゼβ又は他のＤＮＡポリメラーゼで、そしてギャップを塞ぐためにＤＮＡリガーゼで処理される。結果として、潜在的に向上した特性を有するアウトプット鎖を生成するために、インプット鎖間で配列が変化したものが再編成される。これらのアウトプットポリヌクレオチドは、適当なベクター内に直接クローニングされることがあり、あるいはそれらはクローニング前にＰＣＲによって増幅されることがある。生じたクローンは、所望の特性における向上についての選択又はスクリーニングにかけられる。

【０１０１】

別の態様は、GRAMMRによるゾーナル突然変異誘発、すなわち、多数の塩基対合の可能性を有するヌクレオチド類似体を用いる、ミスマッチの残基における、又はその近傍における

ランダムな又はセミランダムな突然変異を提供する。これは、注目の特定の部分において本質的にランダムな突然変異誘発の集中を提供し、そして本発明に別の利益を付加する。若干異なる機能を有する類似の遺伝子、例えば植物のＲ－遺伝子、酵素等は、それら自身の特定の活性にとって重要であろう領域において、それらの間で、中程度の配列の差異を示す。これらの活性、例えば異なる基質、結合の相手、制御部位等を発現する遺伝子は、これらの機能を支配する領域において、不均一性を有するはずである。そのような機能の特異性がこれらのアミノ酸及びそれらに隣接するものと関連していることが知られているので、GRAMMR突然変異誘発は、遺伝子間の配列のばらつきを再編成し、そして更に、潜在的にランダム化に耐性がない他の配列、例えば構造的骨格、不変の残基、及び他のそのような重要部位を乱すことなく、それらをより更に且つより早く進化させるためにこれらの領域にランダムな突然変異誘発を向けさせる役割を果たすことがある。

【０１０２】

異なる機能を有する、異なる酵素は、作業領域、例えば活性部位、制御部位等だけが異なるのではない。それらは、遺伝的浮動を介して生じる、お互いに由来する他の差異を有することがある。そのような変化の部位における更なるランダム化は、それ故に、突然変異誘発の実験結果にとって、中立であるか、最低限重要であるか、又は有害であると考えられる。ランダムな突然変異誘発をそのような重要でない部位から遠ざけ、且つランダムな突然変異誘発にとってよりよい結果を提示しうる部位、例えば酵素の活性部位に向けるために、当該遺伝子のコドンの使用頻度の偏りが、それらの領域におけるヌクレオチドの全体的な相補レベルを低下させ、又は増大させるために操作されうる。より大きな相補性の領域が、あまり相補的でない領域よりもGRAMMRに対して耐性がない場合、所定の部位での、GRAMMRが指定するゾーナルランダム突然変異誘発の程度は調節されうる。

【０１０３】

別の態様において、ヘテロ二本鎖が形成された後、３’→５’活性を有する酵素が、ヘテロ二本鎖の各末端の一方の鎖が後方に消化されるように添加される。平均して、所望の程度の３’→５’の消化が起きた時点で、同一又は追加の酵素に由来する５’→３’ポリメラーゼ活性に、鋳型として逆鎖を用いて二本鎖を回復させるために、dNTPが添加される。このように、消化された領域におけるミスマッチは、相補なものへと解消される。任意に、生じた二本鎖は精製され、変性され、そしてアニーリングさせられる。消化、その後のポリマー化の過程が繰り返されて、新規のキメラ配列が生じる。当該過程の更なるサイクルは、所望により実施されうる。アウトプットの二本鎖分子は、クローニングされ、そして所望の機能的特性について試験される。この過程は、フラグメント化及び再編成を必要としない。更に、この過程はエンドヌクレアーゼによる開裂を必要としない。

【０１０４】

別の態様において、ヘテロ二本鎖分子が形成された後、５’→３’エキソヌクレアーゼ活性を有する酵素、例えば、Enger, MJ and Richardson, CC, J Biol Chem 258 (83) 11197に開示されているようなT7遺伝子6エキソヌクレアーゼが、ヘテロ二本鎖の各末端の一方の鎖が消化されるように添加される。平均して、所望の程度の５’→３’消化が起こった時点で、当該反応は停止され、そしてエキソヌクレアーゼは不活性化される。標的ポリヌクレオチドの５’及び３’末端に相補的なオリゴヌクレオチドプライマーが添加され、そしてアニーリングされる。ＤＮＡポリメラーゼ、例えばT4 DNAポリメラーゼ、ＤＮＡリガーゼ及びdNTPが、５’→３’ポリメラーゼ活性にプライマーを伸長させ、そして鋳型として逆鎖を用いて二本鎖を回復させ、同時にリガーゼにニックを塞がせるために添加される。このように、消化した領域におけるミスマッチは相補的なものへと解消される。任意に、生じた二本鎖は任意に、生じた二本鎖は精製され、変性され、そしてアニーリングさせられる。消化、その後のポリマー化の過程が繰り返されて、新規のキメラ配列が生じる。当該過程の更なるサイクルは、所望により実施されうる。アウトプットの二本鎖分子は、クローニングされ、そして所望の機能的特性について試験される。この過程は、フラグメント化及び再編成を必要としない。更に、この過程はエンドヌクレアーゼによる開裂を必要としない。

【０１０５】
本発明において、ランダムな再編成は、in vitroでのＤＮＡミスマッチ解消反応において起こる。この方法は、初期の突然変異再編成（「シャッフリング」）法の基礎を務める「遺伝子再会合」の段階を全く必要としない。その代わりに、再構成された、又は人工的なＤＮＡのミスマッチを解消する系の、１又は複数のＤＮＡの鎖から別のＤＮＡの鎖へと、ハイブリダイゼーション及びin vitroでのミスマッチの解消によって配列のばらつきを伝達する能力を基にしている。

【０１０６】
通常、組み換えＤＮＡ技術の標準的な技術は、種々の刊行物、例えば引用によって全体が本明細書に組み入れられる、Ausubel, 1987; Ausubel, 1999; Sambrook et al., 1989に記載されている。ポリヌクレオチドを修飾する酵素は、取扱説明書に従い使用された。所望により、推定されるＤＮＡ配列を増幅させるためのＰＣＲアンプリマーが、技術者の裁量で選択されることがある。

【０１０７】
GRAMMR反応に関連して本発明において教示した各活性は、類似の活性を有する機能的に等価な因子と交換されてもよく、そしてそのような変化が本発明の範囲内にあることに注意すること。例えば、例２で指摘したように、Taq DNAリガーゼは、T4 DNAリガーゼの代わりとなることがある。他のリガーゼ、例えばE. コリのＤＮＡリガーゼも代わりとなることがある。同様に、それぞれ例２及び８に示すように、Pfuポリメラーゼ及びT7 DNAポリメラーゼもT4 DNAポリメラーゼの代わりとなることができる。適当なエキソヌクレアーゼ活性を有する他の酵素は、関連するポリメラーゼの有無に関係なく、これらの酵素の代わりに、GRAMMR反応に必要なエキソヌクレアーゼ活性について機能しうる。同様に、GRAMMRのための働くことが示されたものと機能的に等しい活性を有する任意なポリメラーゼも代わりとなることができる。これらには、中でも、E.コリのPol 1、E.コリのPol 1のクレノーフラグメント、ポリメラーゼベータが含まれる。

【０１０８】
鎖の開裂は様々な方法でもたらされうる。CEL Iに加えて、多くの機能的等価物、及び様々な植物種由来の抽出物に見られる潜在的に相同の活性(Oleykowski, Nucleic Acids Res 1998 ; 26: 4597-602)が使用されうる。他のミスマッチに対するエンドヌクレアーゼ、例えばT4エンドヌクレアーゼVII、T7エンドヌクレアーゼI、及びSPヌクレアーゼ(Oleykowski, Biochemistry 1999 ; 38: 2200-5)が使用されうる。一本鎖ＤＮＡを攻撃する他のヌクレアーゼ、例えばS1 ヌクレアーゼ, FEN1, クリーバーゼ, リョクトウヌクレアーゼ, 及びヌクレアーゼ P1も使用されうる。ＤＮＡにおいてランダムに開裂を引き起こす酵素、例えば、膵臓のデオキシリボヌクレアーゼＩも、GRAMMRにおける鎖の開裂活性の代わりとなりうる。他の手段を介して鎖の開裂をもたらす多くの方法が更に予想される。これらには、酢酸テトラエチルアンモニウムと一緒に使用する過マンガン酸カリウム、立体的に嵩高い、光で活性化されうるＤＮＡ干渉因子、例えば[Rh(bpy)2(chrysi)]3+、四酸化オスミウムとピペリジンアルカロイド、及びヒドロキシルアミンとピペリジンアルカロイドの使用、並びに鎖の破壊をもたらす放射エネルギーの使用が含まれる。

【０１０９】
CEL Iは、セロリから単離されたミスマッチエンドヌクレアーゼである。標的ポリヌクレオチド配列、例えばガンと関連するものにおける突然変異の検出のための診断方法におけるCEL Iの使用は、米国特許第5,869,245号に開示されている。CEL Iを単離し、そして調製する方法もこの特許において開示されている。しかしながら、ＤＮＡ配列の再編成におけるCEL Iの使用に関しては、この特許に開示されていない。

【０１１０】
CEL Iをコードする核酸分子は、PCT公開公報WO01/62974 A1に開示されている。米国特許第5,869,245号と同様に、ガンと関連する標的ポリヌクレオチド配列における突然変異の検出のための診断方法におけるCEL Iの使用が開示されている。また同様に、DNAの再編成におけるCEL Iの使用に関しては開示されていない。

【０１１１】
ファージT4のエンドヌクレアーゼVIIと、８，４，又は１個のヌクレオチド、あるいは８個の可能性のある塩基のミスマッチのいずれかのＤＮＡループとのin vitroでの反応性が、「Endonuclease VII of Phage T4 Triggers Mismatch Correction in Vitro」Solaro, et al., J Mol Biol 230 (93) 868において開示されている。当該刊行物は、エンドヌクレアーゼVIIが、誤対合の６個のヌクレオチドの３側にニック及びカウンターニックを作り出すことによって、二本鎖の破壊を導入する機構を報告している。当該刊行物は、最初のニックの発生とカウンターニックの発生との間の時間の遅れが、gp43の３’－５’エキソヌクレアーゼ活性に誤対合を除去させ、そしてそのポリメラーゼ活性にカウンターニックの発生前にギャップを埋めさせるのに十分であったことを開示している。ヌクレオチドは、いずれか一方の鎖上にあるミスマッチの３’側に位置し、そして最初の安定な塩基対にあるミスマッチの５’側で停止する最初のニックから除去される。ポリメラーゼ活性は、最初のニックに対して５’から３’方向に進み、これはＤＮＡリガーゼによって塞がれる。その結果、３～４個のヌクレオチドの非常に短い修復トラック(track)が前のニックに及ぶ。当該刊行物は、種々の活性に関して、エンドヌクレアーゼVIIがファージT4内にありうることを結論づけている。しかしながら、当該刊行物は、ファージT4以外のエンドヌクレアーゼVIIについての実際の有用性について何ら開示しておらず、そしてＤＮＡ再編成におけるその利用可能性に関しても開示していない。

【０１１２】
エスケリッチャ・コリにおけるin vivoでのキメラＤＮＡ配列のライブラリーを作製する方法は、Nucleic Acids Research, 1999, Vol 27, No. 18, el8, Volkov, A. A., Shao, Z., and Arnold, F. Hに開示されている。当該方法は、E.コリを形質転換するためにin vitroで形成したヘテロ二本鎖を使用しており、この中で、ヘテロ二本鎖内の同一でない領域の修復が、それぞれの親の因子から構成される新規な組み換え配列のライブラリーを作り出している。当該刊行物は、既存のＤＮＡ組み換え法、すなわちＤＮＡシャッフリングに対する、簡便な付加的なものとしてのこの方法の使用を開示しているが、開示された方法は、E.コリのin vivoの環境に限定されている。当該刊行物は、E.コリにおいて、ミスマッチ修復について利用可能な機構が複数存在し、そしてMutS/L/Hの酵素系を利用する「ロングパッチ(long patch)」の修復機構がヘテロ二本鎖の修復におそらく必要であることを述べている。

【表１】

引用文献

1. Arkin, A.P. and Youvan, D.C. (1992) An algorithm for protein engineering: simulations of recursive ensemble mutagenesis. *Proc Natl Acad Sci U S A*, **89**, 7811-7815.

2. Ausubel, F.M. (1987) *Current protocols in molecular biology*. Published by Greene Pub. Associates and Wiley-Interscience : J. Wiley, New York.

3. Ausubel, F.M. (1999) *Short protocols in molecular biology : a compendium of methods from Current protocols in molecular biology*. Wiley, New York.

【表２】

4. Barnes, W.M. (1994) PCR amplification of up to 35-kb DNA with high fidelity and high yield from lambda bacteriophage templates. *Proc Natl Acad Sci U S A*, **91**, 2216-2220.

5. Bartel, D.P. and Szostak, J.W. (1993) Isolation of new ribozymes from a large pool of random sequences. *Science*, **261**, 1411-1418.

6. Cadwell, R.C. and Joyce, G.F. (1992) Randomization of genes by PCR mutagenesis. *PCR Methods Appl*, **2**, 28-33.

7. Calogero, S., Bianchi, M.E. and Galizzi, A. (1992) In vivo recombination and the production of hybrid genes. *FEMS Microbiol Lett*, **76**, 41-44.

8. Caren, R., Morkeberg, R. and Khosla, C. (1994) Efficient sampling of protein sequence space for multiple mutants. *Biotechnology (N Y)*, **12**, 517-520.

9. Delagrave, S., Goldman, E.R. and Youvan, D.C. (1993) Recursive ensemble mutagenesis. *Protein Eng*, **6**, 327-331.

10. Delagrave, S. and Youvan, D.C. (1993) Searching sequence space to engineer proteins: exponential ensemble mutagenesis. *Biotechnology (N Y)*, **11**, 1548-1552.

11. Goldman, E.R. and Youvan, D.C. (1992) An algorithmically optimized combinatorial library screened by digital imaging spectroscopy. *Biotechnology (N Y)*, **10**, 1557-1561.

【表３】

12. Gram, H., Marconi, L.A., Barbas, C.F.d., Collet, T.A., Lerner, R.A. and Kang, A.S. (1992) In vitro selection and affinity maturation of antibodies from a naive combinatorial immunoglobulin library. *Proc Natl Acad Sci U S A*, **89**, 3576-3580.

13. Hayashi, N., Welschof, M., Zewe, M., Braunagel, M., Dubel, S., Breitling, F. and Little, M. (1994) Simultaneous mutagenesis of antibody CDR regions by overlap extension and PCR. *Biotechniques*, **17**, 310, 312, 314-315.

14. Hermes, J.D., Blacklow, S.C. and Knowles, J.R. (1990) Searching sequence space by definably random mutagenesis: improving the catalytic potency of an enzyme. *Proc Natl Acad Sci U S A*, **87**, 696-700.

15. Holland, J.H. (1992) *Adaptation in natural and artificial systems : an introductory analysis with applications to biology, control, and artificial intelligence*. MIT Press, Cambridge, Mass.

16. Ji, G. and Silver, S. (1992) Regulation and expression of the arsenic resistance operon from Staphylococcus aureus plasmid pI258. *J Bacteriol*, **174**, 3684-3694.

17. Kauffman, S.A. (1993) *The origins of order : self-organization and selection in evolution*. Oxford University Press, New York.

【表４】

18. Marton, A., Delbecchi, L. and Bourgaux, P. (1991) DNA nicking favors PCR recombination. *Nucleic Acids Res*, **19**, 2423-2426.

19. Meyerhans, A., Vartanian, J.P. and Wain-Hobson, S. (1990) DNA recombination during PCR. *Nucleic Acids Res*, **18**, 1687-1691.

20. Nissim, A., Hoogenboom, H.R., Tomlinson, I.M., Flynn, G., Midgley, C., Lane, D. and Winter, G. (1994) Antibody fragments from a 'single pot' phage display library as immunochemical reagents. *Embo J*, **13**, 692-698.

21. Oleykowski, C.A., Bronson Mullins, C.R., Godwin, A.K. and Yeung, A.T. (1998) Mutation detection using a novel plant endonuclease. *Nucleic Acids Res*, **26**, 4597-4602.

22. Oliphant, A.R., Nussbaum, A.L. and Struhl, K. (1986) Cloning of random-sequence oligodeoxynucleotides. *Gene*, **44**, 177-183.

23. Sambrook, J., Maniatis, T. and Fritsch, E.F. (1989) *Molecular cloning : a laboratory manual*. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.

24. Stemmer, W.P. (1994a) DNA shuffling by random fragmentation and reassembly: in vitro recombination for molecular evolution. *Proc Natl Acad Sci U S A*, **91**, 10747-10751.

【表５】

25. Stemmer, W.P. (1994b) Rapid evolution of a protein in vitro by DNA shuffling. *Nature*, **370**, 389-391.

26. Stemmer, W.P., Morris, S.K. and Wilson, B.S. (1993) Selection of an active single chain Fv antibody from a protein linker library prepared by enzymatic inverse PCR. *Biotechniques*, **14**, 256-265.

27. Winter, G., Griffiths, A.D., Hawkins, R.E. and Hoogenboom, H.R. (1994) Making antibodies by phage display technology. *Annu Rev Immunol*, **12**, 433-455.

28. Yang, B., Wen, X., Kodali, N.S., Oleykowski, C.A., Miller, C.G., Kulinski, J., Besack, D., Yeung, J.A., Kowalski, D. and Yeung, A.T. (2000) Purification, cloning, and characterization of the CEL I nuclease. *Biochemistry*, **39**, 3533-3541.

【０１１３】
次の限定的でない実施例は、本発明を例示するために供される。
【０１１４】
実施例１
CEL Iによるミスマッチの ＤＮＡ基質の開裂
本実施例は、CEL I酵素の調製及びミスマッチの ＤＮＡ基質の開裂におけるその使用を教示する。
【０１１５】
CEL I酵素は、ホモジェナイゼーション、硫酸アンモニウム、及びYang等によって記載されたコンカナバリンＡセファロースプロトコール(Biochemistry, 39:3533-3541 (2000). これは引用によって本明細書に組み入れられる)を用いて、セロリの柄から調製された。冷凍されたセロリの柄の１．５ｋｇの試料をジューサーでホモジェナイズした。１リットルのジュースを回収し、１００マイクロモーラーのフェニルメチルスルホニルフッ化物(PMSF)を含む１００ｍＭのＴｒｉｓ－ＨＣｌで調整し、そして二層のミラクロスを介して濾過した。固体の$(NH_4)_2SO_4$を、氷上で撹拌しながらゆっくりと２５％飽和まで添加した。３０分後、懸濁液を２７，０００ｇ、４℃で１．５時間遠心した。懸濁液を回収し、そして氷上で撹拌しながら$(NH_4)_2SO_4$で８０％飽和に調整し、２７，０００ｇで２時間遠心した。ペレットを緩衝液Ｂ（０．１ＭＴｒｉｓ－ＨＣｌ，ｐＨ７．７，０．５Ｍ ＫＣｌ，１００マイクロモーラーのＰＭＳＦ）中で再懸濁し、そして同一の緩衝液に対して透析した。
【０１１６】
コンカナバリンＡ(ConA)セファロースアフィニティークロマトグラフィーは、透析された試料と２ｍｌのConA樹脂を一晩おだやかに振とうしながら最初のインキュベートをすることによって実施された。ConA樹脂は、続いて０．５ｃｍの直径のカラムに充填され、そしてカラムの体積の緩衝液Ｂで複数回洗浄された。画分を１ｍｌのアリコートに回収した。

画分は、放射性標識したミスマッチの基質に対するミスマッチの開裂活性について、引用によって本明細書に組み入れられるOleykowski et al. (Nucleic Acids Research 26: 4597-4602 (1998)に記載のように、０．１μｌの各画分と緩衝液Ｄ（２０ｍＭＴｒｉｓ－ＨＣｌ，ｐＨ７．４，２５ｍＭ ＫＣｌ，１０ｍＭ $MgCl_2$）中のミスマッチのプローブを４５℃で３０分間インキュベートをすることによってアッセイされた。反応産物は、７％ウレア(Invitrogen)を含む１０％TBE-PAGEゲル上での分離、続くオートラジオグラフィーによって可視化された。続いて、CEL Iの第五画分の一連の５倍希釈したものが、放射性標識したミスマッチ基質のミスマッチ開裂について解析された。反応は、緩衝液 D、New England BioLabs (NEB)のT4 DNA リガーゼ 緩衝液 (50 mM Tris-HCL, pH 7.5,10 mM MgCl2,10 mM ジチオトレイトール (DTT), 1 mM ATP, 25 μｇ/ml BSA), 又はGibco/BRL T4 DNA リガーゼ 緩衝液 (50 mM Tris-HCL, pH 7.6,10mM $MgCl_2$, 1 mM DTT, 1 mM ATP, 5% (w/v) ポリエチレングリコール-8000)のいずれかにおいて実施された。反応産物は上述のように可視化された。緩衝液Ｄ及びNEB T4 DNA リガーゼ緩衝液中での開裂活性は、おおよそ等しいことが明らかとなり、一方、PEG含有Gibco/BRL リガーゼ緩衝液は他の緩衝液と比較して５～１０倍に増強された。

【０１１７】

CEL I活性の更なる解析は、基質として、異なる緑色蛍光タンパク質(GFP)由来の確定している二本鎖ＤＮＡを用いて実施された。このGFPのヘテロ二本鎖基質は、センス鎖上のサイクル３ＧＦＰ及びアンチセンス鎖上の野生型ＧＦＰに相当する一本鎖ＤＮＡをアニーリングすることによって調製した。一本鎖ＤＮＡは、非対称ＰＣＲによって合成され、そしてアガロースゲル電気泳動によって単離された。９０℃に加熱し、そして1xNEB制限酵素緩衝液２２ (10 mM Tris-HCL, pH 7.9,10 mM MgCl2,50 mM NaCl, 1 mM ジチオトレイトール)の存在下で冷却することによってアニーリングした後、ヘテロ二本鎖ＤＮＡを、アガロースゲル電気泳動、続いてヘテロ二本鎖バンドの切り出し及びQiaquick DNAスピンカラムを用いる抽出によって単離した。１又は２個のヌクレオチドの長さの、合計２８個のミスマッチが、ヘテロ二本鎖分子全体で生じる。ミスマッチの分布は、１又は２個のヌクレオチドによって分離された複数のミスマッチの小さな集団から、一方に存在する３０個以上の塩基対によって分離されたミスマッチに及ぶ。

【０１１８】

1XNEB T4 DNA リガーゼ緩衝液中の、一連の３倍希釈したCEL Iが調製され、そしてそれぞれの１μｌのアリコートが、それぞれ基質として０．５μｇの高次コイルのプラスミド調製物又は１００ｎｇのサイクル３／野生型GFPのヘテロ二本鎖を含む、２つの別々の１０μｌの反応液中でインキュベートされた。全ての反応が、1XNEB T4 DNA リガーゼ緩衝液中で行われた。反応液は、４５℃で３０分間インキュベートされ、そして臭化エチジウムの存在下で、１．５％ＴＢＥアガロースゲルにかけられた。

【０１１９】

CEL Iを増量しつつ高次コイルのプラスミド調製物を処理した結果、高次コイルのプラスミドは、ニックのある環、続いて直鎖状の分子、そして次にランダムなサイズのＤＮＡのより小さいフラグメントへと変換された。CEL I調製物によるミスマッチGFP基質の処理は、完全長のヘテロ二本鎖を、ミスマッチの集合の近傍で逆のＤＮＡ鎖上の開裂を示す確率が高いラダーＤＮＡに変換させた。更なる消化は、ミスマッチのGFP基質の、CEL I調製物によるヘテロ二本鎖ＤＮＡの消化の限界を表すと思われるより小さなＤＮＡへの変換をもたらした。

【０１２０】

例２

ミスマッチ解消カクテルによる完全長GFP遺伝子の変換

本実施例は、完全長のGFP遺伝子を保存する、種々のミスマッチ解消カクテルを教示する。

【０１２１】

ミスマッチのGFP基質は、合成のミスマッチ解消システムを共に構成する酵素のカクテル

の存在下、種々の濃度のCEL Idで処理された。使用した酵素は、CEL I,T4 DNAポリメラーゼ，Taq DNAポリメラーゼ及びT4 DNAリガーゼであった。CEL I活性は、ヘテロ二本鎖の、ミスマッチ塩基の３’側に切れ目をいれるはずである。T4 DNAポリメラーゼは、ニックのあるヘテロ二本鎖からのミスマッチ塩基の除去のために３’－５’エキソヌクレアーゼを含む。T4 DNAポリメラーゼ及びTaq DNAポリメラーゼは、ギャップを埋めることができるＤＮＡポリメラーゼを含む。T4 DNAリガーゼは修復された分子のニックを塞ぐ。Taq DNAポリメラーゼは、更に５’フラッパーゼ活性を有する。

【０１２２】

マトリックスの実験は、GFPヘテロ二本鎖基質内のミスマッチを解消させる働きをするであろう反応条件を同定するために実施された。１つの実験において、サイクル３／野生型GFPのヘテロ二本鎖は、８つの異なる濃度のCEL I第五画分（上述）の連続希釈液と一緒に、マトリックスフォーマット内でインキュベートされた。各反応液は、１００ｎｇのヘテロ二本鎖基質及び1x NEBT4 DNAリガーゼ緩衝液中の０．２μｌのT4 DNAリガーゼ(Gibco BRL)及び各２５０μMのdNTPを、１０μｌの反応液量中に含んだ。合計して、マトリックスは９６個の別々の反応液を含んだ。一式の反応液が、室温で３０分間インキュベートされ、一方、他方の一式が３７℃で３０分間インキュベートされた。

【０１２３】

インキュベーション後、ＰＣＲが、各反応からGFP遺伝子を増幅させるために使用された。各ＰＣＲ由来のアリコートが、続いてHindIII及びHpaIで消化され、そして臭化エチジウムを含む３％アガロースゲル上で電気泳動にかけられた。サイクル３のGFPのみがHindIII部位を有しており、野生型のみがHpaI部位をコードしている。

【０１２４】

ＤＮＡのミスマッチ解消がHindIII又はHpaIのミスマッチ部位のいずれかで起こった場合、ある割合のＰＣＲ産物が、両方の部位を含み、新規バンドを生成することが予想される。当該バンドは、CEL I、T4 DNAポリメラーゼ、Taq DNAポリメラーゼのいずれも含まないネガティブコントロール試料を含む全ての試料中で観察された。この結果は、基底レベルのバックグラウンドの組み替えが、GRAMMR反応以外の実験、恐らくはＰＣＲ段階のある時点で起こったようであることを示した。ＰＣＲが媒介する組み換えは、増幅の間に、既知の配列間においてある頻度で起こることが知られている[Paabo, et al.,DNA damage promotes jumping between strands during enzymatic amplification. J Biol Chem 265 (90) 4718-4721参照]。

【０１２５】

別の実験において、２００ｎｇのサイクル３／野生型GFPのヘテロ二本鎖が、種々の濃度のCEL I及びT4 DNAポリメラーゼを、２．５ユニットのTaq DNAポリメラーゼと一緒に用い、T4 DNAリガーゼ（０．２ユニット；Gibco BRL）の存在下又は不在下で処理された。各反応液は、０．０５ｍMの各dNTPを含む1X NEB T4 DNA リガーゼ 緩衝液を２０μｌの最終量中に含んだ。反応液は、３７℃で３０分間インキュベートされ、そして１０μｌが、臭化エチジウムの存在下、２％ＴＢＥアガロースゲル上に流された。結果は、ＤＮＡリガーゼの存在下で、しかしT4 DNAポリメラーゼの不在下で、CEL Iの量を増やすことが、ヘテロ二本鎖ＤＮＡのより大きな分解をもたらし、しかし、この効果が、反応液中のT4 DNAポリメラーゼの量を増大させることによって中和されうることを示した。これらの結果は、完全な反応の種々の構成成分が一緒に、ＤＮＡのミスマッチの解消を介して完全長遺伝子の完全性を保存する働きをしうることを示した。

【０１２６】

別のマトリックス実験は、これらの結果を更に詳しく調べ、そしてこの合成系のＤＮＡミスマッチ解消のための更なる条件を同定するために実施された。６０ｎｇのサイクル３／野生型GFPのヘテロ二本鎖が、２．５ユニットのTaq DNAポリメラーゼ及び０．２ユニットのT4 DNAポリメラーゼの存在下で、０．５ｍMの各dNTPを含む1x NEBT4 DNAリガーゼ緩衝液中で、１０μｌの反応液量で、種々の濃度のCEL I及びT4 DNAポリメラーゼで処理された。各反応液を２０℃、３０℃、３７℃、又は４５℃のいずれかで、１時間インキュベー

トした。全ての反応液が、臭化エチジウムの存在下で、１．５％ＴＢＥアガロースゲルに流された。この結果は、GFPのヘテロ二本鎖が、CEL I調製物のみによって別個のフラグメントに開裂されたことを示した。ＤＮＡミスマッチ解消の成功は、最初に、見かけ上のGFP配列の完全長の完全性が、ミスマッチ解消系の他の構成成分によって、CEL Iの存在下で維持された程度によって評価された。このアッセイにおいて完全長分子として高率のＤＮＡを保存した酵素濃度及び温度の条件が同定された。すなわち、１μｌのCEL I画分の５つの調製物（実施例１に記載）と1μｌ（１ユニット）のT4 DNAポリメラーゼと、当該実験で一定に保たれた他の反応成分の存在下、である。反応温度が増大するにつれ、CEl Iの分解活性がそれに応じて増大したことが明らかとなった。更に、修復反応の他の成分が２０℃、３０℃、及び３７℃で完全長ＤＮＡの完全性を保存するために働くが、４５℃で完全長のＤＮＡを保存するには全く有効ではなかったことを示した。これらの結果から、我々は、これらの実験条件下、４５℃でのインキュベーションがGRAMMRの過程にとって最適でなく、そして２０℃、３０℃、及び３７℃でのインキュベーションが許容されると結論づけた。

【０１２７】

代わりの酵素をＤＮＡミスマッチ解消反応に使用した別の実験が実施された。T4 DNAリガーゼの代わりに、Taq DNAリガーゼが使用された。Pfu DNAポリメラーゼ(Stratagene)が、３’エキソヌクレアーゼ／ポリメラーゼとしてT4 DNAポリメラーゼを含んだ一組の反応との対応する比較において利用された。反応は、８ユニットのTaq DNAリガーゼ(NEB)、２．５ユニットのTaq DNAポリメラーゼ、０．５ｍＭの各dNTP、CEL Iの様々な希釈液及びT4 DNAポリメラーゼ又はPfu DNAポリメラーゼのいずれかを含むTaq DNAリガーゼ緩衝液中で実施された。反応は、臭化エチジウムの存在下、１．５％ＴＢＥアガロースゲルに流された。Pfu DNAポリメラーゼ、Taq DNAポリメラーゼ、及びTaq DNAリガーゼの存在下で、CEL Iで処理した基質ＤＮＡの完全長の完全性が、CEL I単独でインキュベートしたＤＮＡと比較して増大されたことが明らかとなった。この結果は、機能的に等しい活性を有する酵素が、GRAMMR反応内で首尾よく置き換えられうることを示す。

【０１２８】

実施例３

ＤＮＡミスマッチ解消の後のGFPヘテロ二本鎖ＤＮＡへの制限部位の復元（GRAMMR)

この実験は、制限部位の復元を証明することによって、ＤＮＡミスマッチ解消による遺伝子再編成（GRAMMR)の実現可能性を教示する。

【０１２９】

上文で明らかとなった最適の条件（１ｘの反応液は、２．５ユニットのTaq DNAポリメラーゼ及び０．２ユニットのT4 DNAポリメラーゼの存在下で、０．５ｍＭの各dNTPを含む1x NEBT4 DNAリガーゼ緩衝液中で、１０μｌの反応液量で、６０μｇのヘテロ二本鎖ＤＮＡ、１μｌのCEL Iの第五画分（実施例１に記載）、１ユニットのT4 DNAポリメラーゼを含んだ）２０倍にスケールアップしたGRAMMR反応の完全長産物が、ゲルで単離され、そして認識部位がGFPのヘテロ二本鎖のミスマッチを覆い、それによって当該ＤＮＡ内のそれらの部位を制限酵素耐性にするエンドヌクレアーゼによる制限解析にかけられた。使用した酵素は、BamHI，HindIII，HpaI，及びXhoIであった。ネガティブコントロールは未処理のGFPのヘテロ二本鎖から構成された。ポジティブコントロールは、それぞれ、サイクル３又は野生型GFP配列から構成された。全てのコントロールは、ＤＮＡミスマッチ解消反応産物と同一の酵素で消化された。全ての試料が、臭化エチジウムの存在下、２％ＴＢＥアガロースゲルに流された。

【０１３０】

ミスマッチ解消カクテルによる処理後、ある比率のＤＮＡが、BamI及びXhoI制限エンドヌクレアーゼに対する感受性を獲得し、これは、ＤＮＡのミスマッチの解消が起こったことを示している。HpaIで切断された試料は、低レベルの開裂がネガティブコントロールで起こったので解明することはできなかった。HindIII，BamI及びXhoI部位は、GRAMMRで処理した試料中で、異なる程度の開裂を示した。XhoI部位の復元は、BamHI部位よりも広範囲

に及んでおり、続いて、これらはHindIII部位の復元よりも広範囲であった。

【０１３１】

開裂が起こる範囲は、ＤＮＡのミスマッチがその部位で解消している範囲を示すものである。ミスマッチの解消効率の差異は、それらの部位に存在するミスマッチの性質及び密度に関係しているようである。例えば、XhoI部位は３つのミスマッチの集合にかかっており、一方、BamHI部位は２つのミスマッチにかかっており、そしてHindIII部位は１つのミスマッチにかかっている。

【０１３２】

実施例４

GRAMMRによって再編成されたGFP遺伝子

この実施例は、GRAMMRがヘテロ二本鎖内の２つの遺伝子配列間の配列のばらつきを再編成することができ、そして直接的にクローニングされ、又はクローニング前にＰＣＲ増幅されたGRAMMR産物中に有意な差異がないことを示す。

【０１３３】

実施例３のGRAMMR処理したＤＮＡ分子は、続いて、PCR-Blunt II-TOPO(Invitrogen)内へのライゲーションによって直接的にクローニングされるか、又はPCRによって増幅され、そして取扱説明書に従いPCR-Blunt II-TOPO内にライゲーションされ、その後E.コリを形質転換させた。個々のクローンを突き、そして液体培地中で生育した後、ＤＮＡを調製し、そしてGFPのインサート配列を決定した。ネガティブコントロールとして、未処理のGFPヘテロ二本鎖基質が、直接的にクローニングされるか、又はプラスミド内にクローニングされる前にＰＣＲ増幅された。

【０１３４】

GRAMMRにおいて、配列情報の再編成は、一方の鎖から他方の鎖への情報伝達の過程から生じる。これらの情報伝達部位は、組み換えに基づいたＤＮＡシャッフリングで起こる交叉の事象に類似している。しかしながら、これらの再編成実験の結果と関連させるために、GRAMMRのアウトプット配列は、交叉の観点で説明される。GRAMMR処理したGFP遺伝子から誘導された２０個の完全長GFPクローンの配列が解析された。これらのクローンのうちの４個は、GRAMMR処理後に直接的にpZeroBlunt[ref]内にクローニングされたＤＮＡ（ＰＣＲ無し）から誘導された。残りの１６個のクローンは、ＰＣＲ後にクローニングされた。これらの完全長GFP配列の解析は、２０個の配列全てが配列の再編成を経験して、１遺伝子あたり１～１０個の交叉を有することを明らかにした。計９９個の交叉がこの遺伝子の組み合わせにおいて見つけられ、１遺伝子あたり平均約５個の交叉を示した。最初と最後のミスマッチ間の約５９０ヌクレオチドの距離で、１２０塩基対あたりおよそ１個の交叉の全体的な頻度が算出された。この２０個のクローンの集合において、合計７個の点突然変異がＰＣＲプライマー間に位置する配列内で発生し、およそ０．０５％の突然変異の頻度をもたらした。

【０１３５】

GRAMMR処理にかけられなかった３５個のクローンの配列決定をした。これらのコントロールのうち、１４個が直接的なクローニングから誘導され、そして２１個のクローンが、鋳型としてGFPヘテロ二本鎖を用いるＰＣＲ増幅後に得られた。これらの３５個のGRAMMR処理されなかったコントロールのクローンのうち、８個が組み換え体であり、交叉は１～３個におよんでおり、最大で１個の交叉が起きている。合計２５個の点突然変異がＰＣＲプライマー間に位置している配列内で起こっており、およそ０．１％の突然変異の頻度をもたらしている。

【０１３６】

有意な差異は、直接的にクローニングされたか、又はＰＣＲ増幅されたGRAMMR処理産物間で観察されなかった。しかしながら、とりわけGRAMMR処理されなかったコントロールにおいて、組み換え体の頻度は、直接的にクローニングされたＤＮＡよりもＰＣＲ増幅されたＤＮＡにおいてより高かった。このより高い頻度は、一定レベルの組み換え体が「ジャンピングＰＣＲ」." [Paabo, et al., DNA damage promotes jumping between templates d

uring enzymatic amplification. J Biol Chem 265 (90) 4718-4721]によって生じていることが明らかとなっている他のものによって得られた結果と一致している。

【０１３７】

実施例５

GFPプラスミドの、ＤＮＡミスマッチ解消によるプラスミド対プラスミド(plasmid-on-plasmid)の遺伝子再編成(POP GRAMMR)のためのヘテロ二本鎖基質の調製

この実験は、ＤＮＡミスマッチ解消による遺伝子再編成のためのヘテロ二本鎖基質が無傷の環状プラスミドの形態であってもよいことを教示する。サイクル３-GFP及び野生型GFPのヘテロ二本鎖分子が、プラスミド対プラスミド(POP)の形式で調製された。この形式において、GFP配列は、環状二本鎖プラスミドベクター主鎖の観点で再編成された。このことは、GRAMMR反応のアリコートを用いるE,コリの直接的な形質転換による再編成産物の回収を可能にした。その結果、GRAMMR処理したＤＮＡのＰＣＲ増幅も他の追加的な操作も、再編成されたクローンを得るために必要ではなかった。

【０１３８】

野生型GFP（配列番号１）及びサイクル３GFP（配列番号２）を含む、POP-GRAMMR反応のためのミスマッチＤＮＡ基質が生成され、２つのpBluescriptを基にしたプラスミド、pBSWTGFP（配列番号３）及びpBSC3GFP（配列番号４）がそれぞれ生じた。GFPは、pBluescriptポリリンカーのKpnIとRcoRI部位との間に挿入され０、その結果、２つのプラスミド間のわずかな配列の差異が、野生型とサイクル３のGFPとが互いに異なる部位で発生した。両プラスミドは、SapIでプラスミドの主鎖を消化し、ＤＮＡスピンカラムを用いて精製し、混合し、１ｘＰＣＲ緩衝液(Barnes, 1994;PNAS, 91, 2216-2220)で整え、沸騰している水槽で３分間加熱し、そして変性したＤＮＡ鎖をアニーリングさせるために室温までゆっくりと冷却した。これらのＤＮＡの変性及びアニーリングは、二本鎖の混合、親の二本鎖の再形成、及び２つのインプットプラスミドのそれぞれに由来する鎖のアニーリングからのヘテロ二本鎖の形成をもたらした。親の二本鎖は、GRAMMRにとって不所望なものとみなされ、そして一方又は他方の親の二本鎖を切断するが、ヘテロ二本鎖分子を切断しない制限酵素による消化によって除去された。PmlI及びXhoIがこの作業のために選択されたのは、PmlIが野生型のGFP配列のみを切断し、XhoIがサイクル３GFPのみを切断するためである。これらの酵素による処置の後、生成物がアガロースゲル上で分離された。完全長の切断されていないヘテロ二本鎖分子が、アガロースゲルにおいてPmlI及びXhoIで切断されたホモ二本鎖から分離され、そして当該バンドの切り出し及びＤＮＡスピンカラムによる精製によって精製された。

【０１３９】

生じたヘテロ二本鎖分子群は、直鎖ＤＮＡを環状に変換するために、ＤＮＡリガーゼを用いて処理された。アガロースゲルのシフト解析による確認の後、環状の二本鎖GFPヘテロ二本鎖プラスミドが、GRAMMR反応の基質として使用された。生じたクローンの例には、配列番号５，配列番号６，配列番号７，及び配列番号８が含まれる。

【０１４０】

実施例６

ＤＮＡのミスマッチ解消による遺伝子再編成のための例示的な反応の要因

CEL IとT4 DNAポリメラーゼの濃度の比較

GRAMMR反応は、多くの酵素活性の相互作用を包含する。GRAMMR反応に関連する複数の要因が試験され、例えばCEL Iの濃度、T4 DNAポリメラーゼの濃度、反応温度、T7 DNAポリメラーゼによるT4 DNAポリメラーゼの置換、Taq DNAポリメラーゼの存在、及びCEL I酵素の起源、である。３つの異なるCEL Iの濃度対２つのT4 DNAポリメラーゼの濃度の行列は、in vitroでのＤＮＡのミスマッチ解消反応の限界を試験するために設定された。

【０１４１】

上述のように調製された２１ナノグラム(21ng)の環状二本鎖ヘテロ二本鎖プラスミドは、1xNEBリガーゼ緩衝液、0.5mMの各dNTP、1.0ユニットのTaq DNAポリメラーゼ、0.2ユニットのT4 DNAリガーゼ(Gibco/BRL)、1.0ユニット又は0.2ユニットのT4 DNAポリメラーゼ、

及び0.3，0.1，又は0.03μｌのCEL Ｉ調製物（実施例１に記載の第五画分）を含む一連の10μｌの反応液中の基質として使用した。２つのT4 DNAポリメラーゼの濃度と３つのCEL Ｉ濃度の、合計６つの組み合わせを表す６つの反応液を調製し、5μｌごとに等しく分け、そして２０℃又は３７℃でインキュベートした。CEL Ｉを含まず、そして0.2ユニットのT4 DNAポリメラーゼと他の反応成分を含むコントロール反応液を調製し、そして３７℃でインキュベートした。３０分後、各反応液の1μｌのアリコートで、コンピテントなDH5-アルファE.コリを形質転換させ、これをLB ampプレート上にプレーティングした。コロニーをつつき、そして培養した。プラスミドＤＮＡを抽出し、そして制限断片長多型(RFLP)によって、続いてGFPの遺伝子配列の配列解析によって試験された。RFLP解析は、野生型とサイクル３のGFP遺伝子間にある複数の制限酵素認識部位の差異を基にした。RFLPの結果は、CEL Ｉ/T4 DNAポリメラーゼ／温度の行列を通じて、制限部位の再編成、すなわちGRAMMRが起こり、そしてそのような再編成がゼロのCEL Ｉコントロールクローンにおいて起こらなかったことを示した。ＤＮＡ配列の解析は、再編成がCEL Ｉ含有試料全てで起こったことを確認した。配列決定はまた、ゼロのCEL Ｉコントロールが、恐らくはE.コリ又はランダム突然変異に起因する、一方の遺伝子配列から他方のものへと一塩基の変化を有していた、１６個のクローンのうちの１個のクローンを除き、再編成されていなかったことも確認した。複数の例示的なGRAMMR再編成GFPクローンの配列を示す；これらは全て、３７℃でインキュベートされた、0.3μｌのCEL Ｉ調製物及び1.0ユニットのT4 DNAポリメラーゼを含む反応液に由来する。親の野生型及びサイクル３のGFP遺伝子を参考のために最初に示す。

【０１４２】

実施例７

Taq DNAポリメラーゼはＤＮＡミスマッチ解消による遺伝子再編成に必要とされない

この実験は、Taq DNAポリメラーゼが、仮にあったとしても、ＤＮＡミスマッチ解消による遺伝子再編成(GRAMMR)の働きに劇的には寄与せず、又はそれを妨害しないことを教示する。Taq DNAポリメラーゼは、５’フラッパーゼ活性を有することが報告されており、そしてGRAMMRを受けるヘテロ二本鎖ＤＮＡ中の不所望な５’フラップの起こりうる形成及び持続に対する保護手段としてこれまでの例に含められてきた。

【０１４３】

GRAMMR反応は、実施例６のように設定され、ここで、２１ナノグラムの環状二本鎖ヘテロ二本鎖プラスミド基質は、10μｌの反応液中に、1xNEBリガーゼ緩衝液、0.5mMの各dNTP、0.2ユニットのT4 DNAリガーゼ、1.0ユニットのT4 DNAポリメラーゼ、及び1.0μｌのCEL Ｉ調製物（実施例１に記載の第五画分）、及び2.5ユニット、0.5ユニット、又は0ユニットのTaq DNAポリメラーゼを含む。３０分後、各反応液の1μｌのアリコートで、コンピテントなDH5-アルファE.コリを形質転換させ、これをLB ampプレート上にプレーティングした。コロニーをつつき、そして培養した。プラスミドＤＮＡを抽出し、そして制限断片長多型(RFLP)によって、続いてGFPの遺伝子配列の配列解析によって試験された。RFLP解析は、野生型とサイクル３のGFP遺伝子間にある複数の制限酵素認識部位の差異を基にした。RFLPの結果は、CEL Ｉ/T4 DNAポリメラーゼ／温度の行列を通じて、制限部位の再編成、すなわちGRAMMRがGRAMMR反応におけるTaq DNAポリメラーゼの存在下でも不在下でも起こったことを示した。ＤＮＡ配列の解析はこれらの結果を確認した。従って、当該データは、Taq DNAポリメラーゼがGRAMMRに不必要であったことを示す。

【０１４４】

実施例８

ＤＮＡミスマッチ解消による遺伝子再編成のための別の校正ＤＮＡポリメラーゼ

この実験は、ＤＮＡミスマッチ解消による遺伝子再編成がT4 DNAポリメラーゼの使用に限定されず、且つ別のＤＮＡポリメラーゼがその代わりとなりうることを教示する。

【０１４５】

反応は、実施例６のように設定され、ここで、２１ナノグラムの環状二本鎖ヘテロ二本鎖プラスミド基質は、10μｌの反応液中に、1xNEBリガーゼ緩衝液、0.5mMの各dNTP、0.2ユニ

ットのT4 DNAリガーゼ(Gibco/BRL)、10ユニット又は2ユニットのT7 DNAポリメラーゼ、及び1.0μlのCEL I調製物（実施例１に記載の第五画分）、及び2.5ユニットのTaq DNAポリメラーゼを含む。３０分後、各反応液の1μlのアリコートで、コンピテントなDH5-アルファE.コリを形質転換させ、これをLB ampプレート上にプレーティングした。コロニーをつつき、そして培養した。プラスミドＤＮＡが抽出され、そしてRFLPによって、続いてGFPの遺伝子配列の配列解析によって試験された。RFLPの結果は、制限部位の再編成、すなわちGRAMMRがT7 ＤＮＡポリメラーゼを含む各反応液中で起こったことを示した。ＤＮＡ配列の解析はこれらの結果を確認した。従って、当該データは、T7 DNAポリメラーゼがGRAMMRのT4 DNAポリメラーゼの代わりとなりうることを示している。更に、個々の成分及び機能が、同様の結果を得つつ、GRAMMRにおいて広範に代替されることがあることを示している。

【０１４６】

実施例９

GRAMMR反応におけるクローニングされたCEL Iの使用

本実施例は、クローニングされた供給源に由来するCEL Iが、セロリから精製された天然のCEL I酵素の代わりに、ＤＮＡミスマッチ解消による遺伝子再編成において、結果が顕著に変化することなく使用されうることを教示する。

【０１４７】

CEL IのcDNAは、セロリRNAからクローニングされた。当該遺伝子は、TMVウイルスベクター内に挿入され、そして発現された。コンストラクトの転写物は、Nicotiana benthamianaa植物に感染させるために使用された。感染組織を採集し、そしてCEL I酵素を精製した。精製酵素を用いて得られたGRAMMRの結果は、セロリから精製されたCEL Iを用いたものと比較され、そして同様であったことが明らかとなった。

【０１４８】

反応は、10μlの反応液中に、1xNEBリガーゼ緩衝液、0.5mMの各dNTP、0.2ユニットのT4 DNAリガーゼ(Gibco/BRL)、1ユニットのT4 DNAポリメラーゼ、及びセロリから調製された1.0μlのCEL I（実施例１に記載の第五画分）、又はクローニングされた供給源から精製された0.3μlのCEL Iを含む、２１ナノグラムの環状二本鎖ヘテロ二本鎖プラスミド基質を用いて設定された。３０分後、各反応液の1μlのアリコートで、コンピテントなDH5-アルファE.コリを形質転換させ、これをLB ampプレート上にプレーティングした。コロニーをつつき、そして培養した。プラスミドＤＮＡが抽出され、そしてRFLPによって、続いてGFPの遺伝子配列の配列解析によって試験された。RFLPの結果は、制限部位の再編成、すなわちGRAMMRがセロリ由来のCEL I、クローニングされたCEL I含有反応液の両方で起こったことを示した。ＤＮＡ配列の解析はこれらの結果を確認した。従って、当該データは、クローニングされる供給源由来のCEL Iがセロリ由来のCEL Iの代わりにGRAMMRに使用されうることを示している。更に、当該データは、CEL I活性が、セロリからCEL Iを抽出するのに使用される精製段階から生じる偶然の作用ではなく、GRAMMR法の一部であることを示す。

【０１４９】

実施例１０

ウイルスベクターにおけるトバモウイルスの30K遺伝子の分子育種

これまでの実施例において、ＤＮＡミスマッチ解消による遺伝子再編成は、高度に相同性がある配列（例えば、wtGFPとサイクル３GFPは９６％同一である）を再編成するのに有用であると教示してきた。本実施例は、GRAMMRが、更に多様性のある核酸配列、例えばトバモウイルス移動(movement)タンパク質遺伝子をコードする遺伝子を再編成するために使用されうることを教示する。

【０１５０】

約７５％同一である２つのトバモウイルス移動タンパク質(MP)遺伝子のヘテロ二本鎖が生成された。ヘテロ二本鎖基質は、非対称ＰＣＲによって合成された逆鎖の部分的に相補的な一本鎖ＤＮＡ（一方の鎖は、タバコモザイクウイルスU1型種(TMV-U1)（配列番号９）を

コードし、そして他方の鎖はタバコモザイクウイルス(ToMV)由来の移動タンパク質遺伝子（配列番号１０）をコードする）をアニーリングすることによって調製された。２つの部分的に相補的な移動タンパク質遺伝子配列は、それらの末端にあるＤＮＡのアニーリングを促進し、且つＰＣＲ増幅及びクローニングを容易にするために、絶対的に相補的な３３個のヌクレオチドを隣接させた。アニーリング反応は、2.5μgの各一本鎖ＤＮＡを、333 mM NaCl, 33 mM $MgCl_2$, 3.3 mM ジチオトレイトール, 166mM Tris-HCl, pH 7を含む150μlの反応液中で混合し、そして９５℃で１分間インキュベートし、続いて室温までゆっくりと冷却することによって起こった。GRAMMRは、5μlのヘテロ二本鎖基質を、1xNEBリガーゼ緩衝液、0.5mMの各dNTP、0.4ユニットのT4 DNAリガーゼ(Gico/BRL)、2.0ユニットのT4 DNAポリメラーゼ、及びCEL Iを含む20μlの反応液中でインキュベートすることによって実施された。CEL Iは、クローニングされた調製物に由来しており、そして使用した量は、2μlの調製物から、５連続の３倍希釈したものにまで及んだ。CEL Iを含まない第７の調製物が調製され、コントロールの役割を果たした。

【０１５１】

室温で１時間後、ＤＮＡは、StratoprepスピンＤＮＡ精製カラム(Stratagene, LaJolla, CA)を用いて反応液から精製され、そして、２つの配列の隣接しているプライマー結合部位にアニーリングするよう設計されたプライマーを用いるＰＣＲ反応のための鋳型として使用された。各反応に由来するＰＣＲ産物は、Strataprepカラムを用いて精製され、AvrII及びPacIで消化され、そして同様に切断されたpGENEWARE-MP-Avr-Pacの移動タンパク質スロット内にライゲーションされた。このプラスミドは、他の移動タンパク質遺伝子による置換を可能にするために、AvrII及びPacIを隣接している移動タンパク質遺伝子で修飾された完全長のトバモウイルス-GFPクローンを含んでいた。DH5-アルファE.コリの形質転換及びプレーティングの後、コロニーをつつき、培養物を生育し、そしてＤＮＡを抽出した。移動タンパク質のインサートは、両方向からのＤＮＡ配列決定にかけられ、そして当該配列データは、GRAMMR処理した材料に由来するインサートの多くが、TMV-U1とToMV両方の移動タンパク質の遺伝子配列から構成された配列を再編成したことを確認した。複数の例示的なGRAMMR MPクローンのＤＮＡ配列を配列番号１１，配列番号１２，配列番号１３，配列番号１４，及び配列番号１５として示す。

【０１５２】

実施例１１

向上したヒ酸塩解毒細菌を産生するためのGRAMMR再編成

ヒ酸塩の解毒は、硫ヒ鉄鉱を含む金鉱の採掘及び他の用途、例えば環境の改善にとって重要である。ヒ酸塩解毒オペロンを含むプラスミドpGJ103(Ji and Silver, 1992) (Ji, G. and Silver, S., Regulation and expression of the arsenic resistance operon from Staphylococcus aureus plasmid pI258, J. Bacteriol. 174,3684-3694 (1992).引用によって本明細書に組み入れられる)を、サイモン・シルバー博士（イリノイ大学、シカゴ、イリノイ州）から入手する。pUC19内にクローニングされたpI258 arsオペロンを含む、E.コリのTG1含有pGJ103は、LBアンピシリンアガープレート上で4μg/mlのMIC（最小阻害濃度）を有する。arsオペロンは、変異原性ＰＣＲ(mutagenic PCR)によって増幅され[REF]、pUC19内にクローニングされ、そしてE.コリ TG1を形質転換させた。形質転換細胞を、一定範囲のヒ酸ナトリウム濃度(2, 4, 8, 16mM)でプレーティングする。最高のヒ酸塩レベルを有する当該プレート由来のコロニーを選択する。当該コロニーを、適切なヒ酸塩選択ができる混合培地中で生育する。プラスミドＤＮＡを培養液から単離する。当該プラスミドＤＮＡを、pUC19プラスミド主鎖内に一度切れ目を入れる制限エンドヌクレアーゼによる消化によって直鎖化する。直鎖化されたプラスミドは、９４℃で１０分間加熱することによって変性される。反応は、一本鎖のアニーリングを促進するために冷却させられる。ハイブリダイズする部分的に相補的な鎖は、ミスマッチの部位に塩基対合していないヌクレオチドを有する。CEL Ｉ（実施例９の方法によって精製）による処理は、各ミスマッチの一方又は他方のポリヌクレオチド鎖の３’側にニックを入れる。３’→５’のエキソヌクレアーゼ（「校正」）活性を含むポリメラーゼ、例えばT4 DNAポリメラーゼの存在は

、ミスマッチの切除を可能にし、そして次に、５’→３’ポリメラーゼ活性は、鋳型として他の鎖を用いてギャップを埋める。続いて、T4 DNAリガーゼが、修復された鎖のリン酸主鎖を修復することによって、ニックを塞ぐ。この結果は、インプットの鎖間の突然変異のランダム化によるものであり、潜在的に向上した特性を有するアウトプット鎖を生成する。これらのアウトプットポリヌクレオチドで、直接的にE.コリ TG1を形質転換し、そしてより高度なヒ酸塩レベル；8，16，32，64mMでプレーティングする。最高のヒ酸塩レベルを有するコロニーを当該プレートから選び、新たな再編成が、生じた形質転換細胞が32，64，128，256mMのヒ酸塩でプレーティングされることを除き、上述の通り実施される。当該工程は、更なる向上を得る目的で、選択されたクローンを用いて１又は複数回繰り返されてもよい。

【図面の簡単な説明】

【０１５３】

【図１】図１は、ミスマッチ解消による遺伝子再編成(GRAMMR)の過程を表す。再編成は、少なくとも２つのヌクレオチドの位置で異なる２つの仮定のポリヌクレオチド間で予期される。２つの位置でミスマッチをもたらす、Ａの上側の鎖及びＢの下側の鎖を示す。再編成ミスマッチ解消の工程の後、４つの異なる産物であるポリヌクレオチド、親の型Ａ及びＢ、並びに再編成産物Ｘ及びＹが見られる。

【図２Ａ】図２は、２つの分子の例示的な部分的に相補的な核酸群を表す。図２Ａは、それぞれ、完全に相補的な上側／下側の鎖１＋／２－及び３＋／４－を有する２つの核酸分子「Ｘ」及び「Ｙ」の配列を示す。核酸ＸとＹの間で異なるヌクレオチドの位置を（★）で示す。

【図２Ｂ】図２Ｂは、変性及びアニーリング後の、核酸Ｘ及びＹに由来する一本鎖の可能性のある組み合わせを示し、そしてこれらの組み合わせのうちのどれが、２つの部分的に相補的な核酸分子群を含んで成るかを示す。

【図２Ｂ】

FIG. 2B

可能性のある鎖の組み合わせ	可能性のある+/-の鎖の組み合わせ	部分的に相補的な群
1+/2-	✓	
1+/3+		
1+/4-	✓	✓
2-/3+	✓	✓
2-/4-		
3+/4-	✓	

FIGURE 2

【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
10 October 2002 (10.10.2002)

PCT

(10) International Publication Number
WO 02/079468 A2

(51) International Patent Classification[7]: C12N 15/10

(21) International Application Number: PCT/US02/03055

(22) International Filing Date: 1 February 2002 (01.02.2002)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/266,386 2 February 2001 (02.02.2001) US
60/268,785 14 February 2001 (14.02.2001) US

(63) Related by continuation (CON) or continuation-in-part (CIP) to earlier applications:
US 60/266,386 (CIP)
Filed on 2 February 2001 (02.02.2001)
US 60/268,785 (CIP)
Filed on 14 February 2001 (14.02.2001)

(71) Applicant (*for all designated States except US*): **LARGE SCALE BIOLOGY CORPORATION** [US/US]; 3333 Vaca Valley Parkway, Suite 1000, Vacaville, CA 95688 (US).

(72) Inventors; and

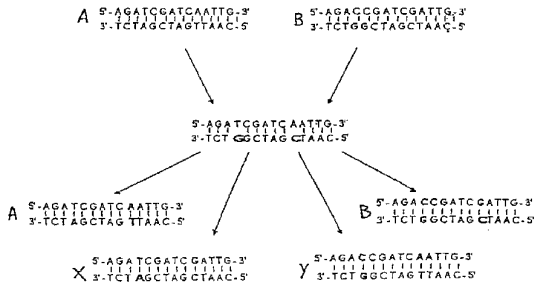
(75) Inventors/Applicants (*for US only*): **PADGETT, Hal, S.** [US/US]; 638 Blossom Oak Court, Vacaville, CA 95687 (US). **FITZMAURICE, Wayne, P.** [US/US]; 1218 Las Encinas Court, Vacaville, CA 95688 (US). **LINDO, John, A.** [US/US]; 143 Sundance Drive, Vacaville, CA 95688 (US).

(74) Agents: **WONG, James, J.** et al.; Campbell & Flores LLP, 4370 La jolla Village Drive, 7th Floor, San Diego, CA 92122 (US).

(81) Designated States (*national*): AE, AG, AL, AM, AT, AT (utility model), AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, CZ (utility model), DE, DE (utility model), DK, DK (utility model), DM, DZ, EC, EE, EE (utility model), ES, FI, FI (utility model), GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU,

[Continued on next page]

(54) Title: A METHOD OF INCREASING COMPLEMENTARITY IN A HETERODUPLEX



(57) Abstract: We describe here an *in vitro* method of increasing complementarity in a heteroduplex polynucleotide sequence. The method uses annealing of opposite strands to form a polynucleotide duplex with mismatches. The heteroduplex polynucleotide is combined with an effective amount of enzymes having strand cleavage activity, 3' to 5' exonuclease activity, and polymerase activity, and allowing sufficient time for the percentage of complementarity to be increased within the heteroduplex. Not all heteroduplex polynucleotide will necessarily have all mismatches resolved to complementary. The resulting polynucleotide is optionally ligated. Several variant polynucleotide result. At sites where either of the opposite strands has templated recoding in the other strand, the resulting percent complementarity of the heteroduplex polynucleotide sequence is increased. The parent polynucleotides need not be cleaved into fragments prior to annealing heterologous strands. Therefore, no reassembly is required.

WO 02/079468 A2

WO 02/079468 A2

SD, SE, SG, SI, SK, SK (utility model), SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZM, ZW.

(84) **Designated States** *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, 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.

A METHOD OF INCREASING COMPLEMENTARITY IN A HETERODUPLEX

BACKGROUND OF THE INVENTION

FIELD OF THE INVENTION

The invention relates generally to molecular biology and more specifically to methods of generating populations of related nucleic acid molecules.

BACKGROUND INFORMATION

DNA shuffling is a powerful tool for obtaining recombinants between two or more DNA sequences to evolve them in an accelerated manner. The parental, or input, DNAs for the process of DNA shuffling are typically mutants or variants of a given gene that have some improved character over the wild-type. The products of DNA shuffling represent a pool of essentially random reassortments of gene sequences from the parental DNAs that can then be analyzed for additive or synergistic effects resulting from new sequence combinations.

Recursive sequence reassortment is analogous to an evolutionary process where only variants with suitable properties are allowed to contribute their genetic material to the production of the next generation. Optimized variants are generated through DNA shuffling-mediated sequence reassortment followed by testing for incremental improvements in performance. Additional cycles of reassortment and testing lead to the generation of genes that contain new combinations of the genetic improvements identified in previous rounds of the process. Reassorting and combining beneficial genetic changes allows an optimized sequence to arise without having to individually generate and screen all possible sequence combinations.

This differs sharply from random mutagenesis, where subsequent improvements to an already improved sequence result largely from serendipity. For example, in order to obtain a protein that has a desired set of enhanced properties, it may be necessary to identify a mutant that contains a combination of various beneficial mutations. If no process is available for combining these beneficial genetic changes, further random mutagenesis will be required. However, random mutagenesis requires repeated cycles of generating and screening large numbers of mutants, resulting in a process that is tedious and highly labor intensive. Moreover, the rate at which sequences incur mutations with undesirable effects increases with the information content of a sequence. Hence, as the information content, library size, and mutagenesis rate increase, the

ratio of deleterious mutations to beneficial mutations will increase, increasingly masking the selection of further improvements. Lastly, some computer simulations have suggested that point mutagenesis alone may often be too gradual to allow the large-scale block changes that are required for continued and dramatic sequence evolution.

There are a number of different techniques used for random mutagenesis. For example, one method utilizes error-prone polymerase chain reaction (PCR) for creating mutant genes in a library format, (Cadwell and Joyce, 1992; Gram et al., 1992). Another method is cassette mutagenesis (Arkin and Youvan, 1992; Delagrave et al., 1993; Delagrave and Youvan, 1993; Goldman and Youvan, 1992; Hermes et al., 1990; Oliphant et al., 1986; Stemmer et al., 1993) in which the specific region to be optimized is replaced with a synthetically mutagenized oligonucleotide.

Error-prone PCR uses low-fidelity polymerization conditions to introduce a low level of point mutations randomly over a sequence. A limitation to this method, however, is that published error-prone PCR protocols suffer from a low processivity of the polymerase, making this approach inefficient at producing random mutagenesis in an average-sized gene.

In oligonucleotide-directed random mutagenesis, a short sequence is replaced with a synthetically mutagenized oligonucleotide. To generate combinations of

distant mutations, different sites must be addressed simultaneously by different oligonucleotides. The limited library size that is obtained in this way, relative to the library size required to saturate all sites, means that many rounds of selection are required for optimization. Mutagenesis with synthetic oligonucleotides requires sequencing of individual clones after each selection round followed by grouping them into families, arbitrarily choosing a single family, and reducing it to a consensus motif. Such a motif is resynthesized and reinserted into a single gene followed by additional selection. This step creates a statistical bottleneck, is labor intensive, and is not practical for many rounds of mutagenesis.

For these reasons, error-prone PCR and oligonucleotide-directed mutagenesis can be used for mutagenesis protocols that require relatively few cycles of sequence alteration, such as for sequence fine-tuning, but are limited in their usefulness for procedures requiring numerous mutagenesis and selection cycles, especially on large gene sequences.

As discussed above, prior methods for producing improved gene products from randomly mutated genes are of limited utility. One recognized method for producing a wide variety of randomly reassorted gene sequences uses enzymes to cleave a long nucleotide chain into shorter pieces. The cleaving agents are then separated from the genetic material, and the material is amplified in such a manner that the genetic material is allowed to reassemble

as chains of polynucleotides, where their reassembly is either random or according to a specific order. ((Stemmer, 1994a; Stemmer, 1994b), US Patent No. 5,605,793, US Patent No. 5,811,238, US Patent No. 5,830,721, US Patent No. 5,928,905, US Patent No. 6,096,548, US Patent No. 6,117,679, US Patent No. 6,165,793, US Patent No. 6,153,410). A variation of this method uses primers and limited polymerase extensions to generate the fragments prior to reassembly (US Patent No. 5,965,408, US Patent No. 6,159,687).

However, both methods have limitations. These methods suffer from being technically complex. This limits the applicability of these methods to facilities that have sufficiently experienced staffs. In addition there are complications that arise from the reassembly of molecules from fragments, including unintended mutagenesis and the increasing difficulty of the reassembly of large target molecules of increasing size, which limits the utility of these methods for reassembling long polynucleotide strands.

Another limitation of these methods of fragmentation and reassembly-based gene shuffling is encountered when the parental template polynucleotides are increasingly heterogeneous. In the annealing step of those processes, the small polynucleotide fragments depend upon stabilizing forces that result from base-pairing interactions to anneal properly. As the small regions of annealing have limited stabilizing forces due to their short length, annealing of highly complementary

sequences is favored over more divergent sequences. In such instances these methods have a strong tendency to regenerate the parental template polynucleotides due to annealing of complementary single-strands from a particular parental template. Therefore, the parental templates essentially reassemble themselves creating a background of unchanged polynucleotides in the library that increases the difficulty of detecting recombinant molecules. This problem becomes increasingly severe as the parental templates become more heterogeneous, that is, as the percentage of sequence identity between the parental templates decreases. This outcome was demonstrated by Kikuchi, *et al.*, (Gene 243:133-137, 2000) who attempted to generate recombinants between *xylE* and *nahH* using the methods of family shuffling reported by Patten *et al.*, 1997; Crameri *et al.*, 1998; Harayama, 1998; Kumamaru *et al.*, 1998; Chang *et al.*, 1999; Hansson *et al.*, 1999). Kikuchi, *et al.*, found that essentially no recombinants (<1%) were generated. They also disclosed a method to improve the formation of chimeric genes by fragmentation and reassembly of single-stranded DNAs. Using this method, they obtained chimeric genes at a rate of 14 percent, with the other 86 percent being parental sequences.

The characteristic of low-efficiency recovery of recombinants limits the utility of these methods for generating novel polynucleotides from parental templates with a lower percentage of sequence identity, that is, parental templates that are more diverse. Accordingly,

there is a need for a method of generating gene sequences that addresses these needs.

The present invention provides a method that satisfies the aforementioned needs, and also provides related advantages as well.

Summary of the Invention

The present invention provides a method for reassorting mutations among related polynucleotides, *in vitro*, by forming heteroduplex molecules and then addressing the mismatches such that sequence information at sites of mismatch is transferred from one strand to the other. In one preferred embodiment, the mismatches are addressed by incubating the heteroduplex molecules in a reaction containing a mismatch nicking enzyme, a polymerase with a 3' to 5' proofreading activity in the presence of dNTPs, and a ligase. These respective activities act in concert such that, at a given site of mismatch, the heteroduplex is nicked, unpaired bases are excised then replaced using the opposite strand as a template, and nicks are sealed. Output polynucleotides are amplified before cloning, or cloned directly and tested for improved properties. Additional cycles of mismatch resolution reassortment and testing lead to further improvement.

Brief Description of the Figures

Figure 1 depicts the process of Genetic ReAssortment by Mismatch Resolution (GRAMMR). Reassortment is contemplated between two hypothetical polynucleotides differing at at least two nucleotide positions. Annealing between the top strand of A and the bottom strand of B is shown which results in mismatches at the two positions. After the process of reassortment mismatch resolution, four distinct product polynucleotides are seen, the parental types A and B, and the reassorted products X and Y.

Figure 2 depicts an exemplary partially complementary nucleic acid population of two molecules. Figure 2A shows the sequence of two nucleic acid molecules "X" and "Y" having completely complementary top/bottom strands 1+/2- and 3+/4-, respectively. The positions of differing nucleotides between the nucleic acids X and Y are indicated (*). Figure 2B shows possible combinations of single strands derived from nucleic acids X and Y after denaturing and annealing and indicates which of those combinations would comprise a partially complementary nucleic acid population of two.

Definitions

As used herein the term "amplification" refers to a process where the number of copies of a polynucleotide is increased.

As used herein, "annealing" refers to the formation of at least partially double stranded nucleic acid by hybridization of at least partially complementary nucleotide sequences. A partially double stranded nucleic acid can be due to the hybridization of a smaller nucleic acid strand to a longer nucleic acid strand, where the smaller nucleic acid is 100% identical to a portion of the larger nucleic acid. A partially double stranded nucleic acid can also be due to the hybridization of two nucleic acid strands that do not share 100% identity but have sufficient homology to hybridize under a particular set of hybridization conditions.

As used herein, "clamp" refers to a unique nucleotide sequence added to one end of a polynucleotide, such as by incorporation of the clamp sequence into a PCR primer. The clamp sequences are intended to allow amplification only of polynucleotides that arise from hybridization of strands from different parents (*i.e.*, heteroduplex molecules) thereby ensuring the production of full-length hybrid products as described previously (Skarfstad, J. Bact, vol 182, No 11, P. 3008-3016).

As used herein the term "cleaving" means digesting the polynucleotide with enzymes or otherwise breaking phosphodiester bonds within the polynucleotide.

As used herein the term "complementary basepair" refers to the correspondence of DNA (or RNA) bases in the double helix such that adenine in one strand is opposite thymine (or uracil) in the other strand and cytosine in one strand is opposite guanine in the other.

As used herein the term "complementary to" is used herein to mean that the complementary sequence is identical to the reverse-complement of all or a portion of a reference polynucleotide sequence or that each nucleotide in one strand is able to form a base-pair with a nucleotide, or analog thereof in the opposite strand. For illustration, the nucleotide sequence "TATAC" is complementary to a reference sequence "GTATA".

As used herein, "denaturing" or "denatured," when used in reference to nucleic acids, refers to the conversion of a double stranded nucleic acid to a single stranded nucleic acid. Methods of denaturing double stranded nucleic acids are well known to those skilled in the art, and include, for example, addition of agents that destabilize base-pairing, increasing temperature, decreasing salt, or combinations thereof. These factors are applied according to the complementarity of the strands, that is, whether the strands are 100% complementary or have one or more non-complementary nucleotides.

As used herein the term "desired functional property" means a phenotypic property, which include but are not limited to, encoding a polypeptide, promoting transcription of linked polynucleotides, binding a protein, improving the function of a viral vector, and the like, which can be selected or screened for. Polynucleotides with such desired functional properties, can be used in a number of ways, which include but are not limted to expression from a suitable plant, animal, fungal, yeast, or bacterial expression vector, integration to form a transgenic plant, animal or microorganism, expression of a ribozyme, and the like.

As used herein the term "DNA shuffling" is used herein to indicate recombination between substantially homologous but non-identical sequences.

As used herein, the term "effective amount" refers to the amount of an agent necessary for the agent to provide its desired activity. For the present invention, this determination is well within the knowledge of those of ordinary skill in the art.

As used herein the term "exonuclease" refers to an enzyme that cleaves nucleotides one at a time from an end of a polynucleotide chain, that is, an enzyme that hydrolyzes phosphodiester bonds from either the 3' or 5' terminus of a polynucleotide molecule. Such exonucleases, include but are not limited to T4 DNA polymerase, T7 DNA polymerase, *E. coli* Pol 1, and Pfu DNA

polymerase. The term "exonuclease activity" refers to the activity associated with an exonuclease. An exonuclease that hydrolyzes in a 3' to 5' direction is said to have "3' to 5' exonuclease activity." Similarly an exonuclease with 5' to 3' activity is said to have "5' to 3' exonuclease activity." It is noted that some exonucleases are known to have both 3' to 5', 5' to 3' activity, such as, *E.coli* Pol I.

As used herein, "Genetic Reassortment by Mismatch Resolution (GRAMMR)" refers to a method for reassorting sequence variations among related polynucleotides by forming heteroduplex molecules and then addressing the mismatches such that information is transferred from one strand to the other.

As used herein, "granularity" refers to the amount of a nucleic acid's sequence information that is transferred as a contiguous sequence from a template polynucleotide strand to a second polynucleotide strand. As used herein, "template sequence" refers to a first single stranded polynucleotide sequence that is partially complementary to a second polynucleotide sequence such that treatment by GRAMMR results in transfer of genetic information from the template strand to the second strand.

The larger the units of sequence information transferred from a template strand, the higher the granularity. The smaller the blocks of sequence information transferred from the template strand, the

lower or finer the granularity. Lower granularity indicates that a DNA shuffling or reassortment method is able to transfer smaller discrete blocks of genetic information from the template strand to the second strand. The advantage of a DNA shuffling or reassortment method with lower granularity is that it is able to resolve smaller nucleic acid sequences from others, and to transfer the sequence information. DNA shuffling or reassortment methods that return primarily high granularity are not readily able to resolve smaller nucleic acid sequences from others.

As used herein the term "heteroduplex polynucleotide" refers to a double helix polynucleotide formed by annealing single strands, typically separate strands, where the strands are non-identical. A heteroduplex polynucleotide may have unpaired regions existing as single strand loops or bubbles. A heteroduplex polynucleotide region can also be formed by one single-strand polynucleotide wherein partial self-complementarity allows the formation of a stem-loop structure where the annealing portion of the strand is non-identical.

As used herein the term "heteroduplex DNA" refers to a DNA double helix formed by annealing single strands, typically separate strands), where the strands are non-identical. A heteroduplex DNA may have unpaired regions existing as single strand loops or bubbles. A heteroduplex DNA region can also be formed by one single-strand polynucleotide wherein partial self-

complementarity allows the formation of a stem-loop structure where the annealing portion of the strand is non-identical.

As used herein the term "homologous" means that one single-stranded nucleic acid sequence may hybridize to an at least partially complementary single-stranded nucleic acid sequence. The degree of hybridization may depend on a number of factors including the amount of identity between the sequences and the hybridization conditions such as temperature and salt concentrations as discussed later.

Nucleic acids are "homologous" when they are derived, naturally or artificially, from a common ancestor sequence. During natural evolution, this occurs when two or more descendent sequences diverge from a parent sequence over time, *i.e.*, due to mutation and natural selection. Under artificial conditions, divergence occurs, *e.g.*, in one of two basic ways. First, a given sequence can be artificially recombined with another sequence, as occurs, *e.g.*, during typical cloning, to produce a descendent nucleic acid, or a given sequence can be chemically modified, or otherwise manipulated to modify the resulting molecule. Alternatively, a nucleic acid can be synthesized *de novo*, by synthesizing a nucleic acid that varies in sequence from a selected parental nucleic acid sequence. When there is no explicit knowledge about the ancestry of two nucleic acids, homology is typically inferred by sequence comparison between two sequences. Where two nucleic acid

sequences show sequence similarity over a significant portion of each of the nucleic acids, it is inferred that the two nucleic acids share a common ancestor. The precise level of sequence similarity that establishes homology varies in the art depending on a variety of factors.

For purposes of this disclosure, two nucleic acids are considered homologous where they share sufficient sequence identity to allow GRAMMR-mediated information transfer to occur between the two nucleic acid molecules.

As used herein the term "identical" or "identity" means that two nucleic acid sequences have the same sequence or a complementary sequence. Thus, "areas of identity" means that regions or areas of a polynucleotide or the overall polynucleotide are identical or complementary to areas of another polynucleotide.

As used herein the term "increase in percent complementarity" means that the percentage of complementary base-pairs in a heteroduplex molecule is made larger.

As used herein the term, "ligase" refers to an enzyme that rejoins a broken phosphodiester bond in a nucleic acid.

As used herein the term "mismatch" refers to a base-pair that is unable to form normal base-pairing interactions (*i.e.*, other than "A" with "T" (or "U"), or "G" with "C").

As used herein the term "mismatch resolution" refers to the conversion of a mismatched base-pair into a complementary base-pair.

As used herein the term "mutations" means changes in the sequence of a wild-type or reference nucleic acid sequence or changes in the sequence of a polypeptide. Such mutations can be point mutations such as transitions or transversions. The mutations can be deletions, insertions or duplications.

As used herein the term "nick translation" refers to the property of a polymerase where the combination of a 5'-to-3' exonuclease activity with a 5'-to-3' polymerase activity allows the location of a single-strand break in a double-stranded polynucleotide (a "nick") to move in the 5'-to-3' direction.

As used herein, the term "nucleic acid" or "nucleic acid molecule" means a polynucleotide such as deoxyribonucleic acid (DNA) or ribonucleic acid (RNA) and encompasses single-stranded and double-stranded nucleic acid as well as an oligonucleotide. Nucleic acids useful in the invention include genomic DNA, cDNA, mRNA and synthetic oligonucleotides, and can represent the sense strand, the anti-sense strand, or both. A nucleic acid

generally incorporates the four naturally occurring nucleotides adenine, guanine, cytosine, and thymidine/uridine. An invention nucleic acid can also incorporate other naturally occurring or non-naturally occurring nucleotides, including derivatives thereof, so long as the nucleotide derivatives can be incorporated into a polynucleotide by a polymerase at an efficiency sufficient to generate a desired polynucleotide product.

As used herein, a "parental nucleic acid" refers to a double stranded nucleic acid having a sequence that is 100% identical to an original single stranded nucleic acid in a starting population of partially complementary nucleic acids. Parental nucleic acids would include, for example in the illustration of Figure 2, nucleic acids X and Y if partially complementary nucleic acid combinations 1+/4- or 2-/3+ were used as a starting population in an invention method.

As used herein, "partially complementary" refers to a nucleic acid having a substantially complementary sequence to another nucleic acid but that differs from the other nucleic acid by at least two or more nucleotides. As used herein, "partially complementary nucleic acid population" refers to a population of nucleic acids comprising nucleic acids having substantially complementary sequences but no nucleic acids having an exact complementary sequence for any other member of the population. As used herein, any member of a partially complementary nucleic acid

population differs from another nucleic acid of the population, or the complement thereto, by two or more nucleotides. As such, a partially complementary nucleic acid specifically excludes a population containing sequences that are exactly complementary, that is, a complementary sequence that has 100% complementarity. Therefore, each member of such a partially complementary nucleic acid population differs from other members of the population by two or more nucleotides, including both strands. One strand is designated the top strand, and its complement is designated the bottom strand. As used herein, "top" strand refers to a polynucleotide read in the 5' to 3' direction and the "bottom" its complement. It is understood that, while a sequence is referred to as bottom or top strand, such a designation is intended to distinguish complementary strands since, in solution, there is no orientation that fixes a strand as a top or bottom strand.

For example, a population containing two nucleic acid members can be derived from two double stranded nucleic acids, with a potential of using any of the four strands to generate a single stranded partially complementary nucleic acid population. An example of potential combinations of strands of two nucleic acids that can be used to obtain a partially complementary nucleic acid population of the invention is shown in Figure 2. The two nucleic acid sequences that are potential members of a partially complementary nucleic acid population are designated "X" (AGATCAATTG; SEQ ID NO:1) and "Y" (AGACCGATTG; SEQ ID NO:2)(Figure 2A). The

nucleic acid sequences differ at two positions (positions 4 and 6indicated by "*"). The "top" strand of nucleic acids X and Y are designated "1+" and "3+," respectively, and the "bottom" strand of nucleic acids X and Y are designated "2-" and "4-," respectively.

Figure 2B shows the possible combinations of the four nucleic acid strands. Of the six possible strand combinations, only the combination of 1+/2-, 1+/4-, 2-/3+, or 3+/4- comprise the required top and bottom strand of a partially complementary nucleic acid population. Of these top/bottom sequence combinations, only 1+/4- or 2-/3+ comprise an example of a partially complementary nucleic acid population of two different molecules because only these combinations have complementary sequences that differ by at least one nucleotide. The remaining combinations, 1+/2- and 2+/4-, contain exactly complementary sequences and therefore do not comprise a partially complementary nucleic acid population of the invention.

In the above described example of a population of two different molecules, a partially complementary population of nucleic acid molecules excluded combinations of strands that differ by one or more nucleotides but which are the same sense, for example, 1+/3+ or 2-/4-. However, it is understood that such a combination of same stranded nucleic acids can be included in a larger population, so long as the population contains at least one bottom strand and at least one top strand. For example, if a third nucleic

acid "2," with strands 5+ and 6- is included, the combinations 1+/3+/6- or 2-/4-/5+ would comprise a partially complementary nucleic acid population. Similarly, any number of nucleic acids and their corresponding top and bottom strands can be combined to generate a partially complementary nucleic acid population of the invention so long as the population contains at least one top strand and at least one bottom strand and so long as the population contains no members that are the exact complement.

The populations of nucleic acids of the invention can be about 3 or more, about 4 or more, about 5 or more, about 6 or more, about 7 or more, about 8 or more, about 9 or more, about 10 or more, about 12 or more, about 15 or more, about 20 or more, about 25 or more about 30 or more, about 40 or more, about 50 or more, about 75 or more, about 100 or more, about 150 or more, about 200 or more, about 250 or more, about 300 or more, about 350 or more, about 400 or more, about 450 or more, about 500 or more, or even about 1000 or more different nucleic acid molecules. A population can also contain about 2000 or more, about 5000 or more, about $1x10^4$ or more, about $1x10^5$ or more, about $1x10^6$ or more, about $1x10^7$ or more, or even about $1x10^8$ or more different nucleic acids. One skilled in the art can readily determine a desirable population to include in invention methods depending on the nature of the desired reassortment experiment outcome and the available screening methods, as disclosed herein.

As used herein, a "polymerase" refers to an enzyme that catalyzes the formation of polymers of nucleotides, that is, polynucleotides. A polymerase useful in the invention can be derived from any organism or source, including animal, plant, bacterial and viral polymerases. A polymerase can be a DNA polymerase, RNA polymerase, or a reverse transcriptase capable of transcribing RNA into DNA.

As used herein the term "proofreading" describes the property of an enzyme where a nucleotide, such as, a mismatch nucleotide, can be removed by a 3'-to-5' exonuclease activity and replaced by, typically, a base-paired nucleotide.

As used herein, a "recombinant" polynucleotide refers to a polynucleotide that comprises sequence information from at least two different polynucleotides.

As used herein the term "related polynucleotides" means that regions or areas of the polynucleotides are identical and regions or areas of the polynucleotides are non-identical.

As used herein the term DNA "reassortment" is used herein to indicate a redistribution of sequence variations between substantially homologous but non-identical sequences.

As used herein the term "replicon" refers to a genetic unit of replication including a length of polynucleotide and its site for initiation of replication.

As used herein the term "sequence diversity" refers to the abundance of non-identical polynucleotides. The term "increasing sequence diversity in a population" means to increase the abundance of non-identical polynucleotides in a population.

As used herein the term "sequence variant" is used herein refers to a molecule (DNA, RNA polypeptide, and the like) with one or more sequence differences compared to a reference molecule. For example, the sum of the separate independent mismatch resolution events that occur throughout the heteroduplex molecule during the GRAMMR process results in reassortment of sequence information throughout that molecule. The sequence information will reassort in a variety of combinations to generate a complex library of "sequence variants".

As used herein the term "strand cleavage activity" or "cleavage" refers to the breaking of a phosphodiester bond in the backbone of the polynucleotide strand, as in forming a nick. Strand cleavage activity can be provided by an enzymatic agent, such agents include, but are not limited to CEL I, T4 endonuclease VII, T7 endonuclease I, S1 nuclease, BAL-31 nuclease, FEN1, cleavase, pancreatic DNase I, SP nuclease, mung bean nuclease, and nuclease P1; by a chemical agent, such

agents include, but are not limited to potassium permanganate, tetraethylammonium acetate, sterically bulky photoactivatable DNA intercalators, [Rh(bpy)2(chrysi)]3+, osmium tetroxide with piperidine, and hydroxylamine with piperidine; or by energy in the form of ionizing radiation, or kinetic radiation.

As used herein the term "sufficient time" refers to the period time necessary for a reaction or process to render a desired product. For the present invention, the determination of sufficient time is well within the knowledge of those of ordinary skill in the art. It is noted that "sufficient time" can vary widely, depending on the desires of the practitioner, without impacting on the functionality of the reaction, or the quality of the desired product.

As used herein the term "wild-type" means that a nucleic acid fragment does not contain any mutations. A "wild-type" protein means that the protein will be active at a level of activity found in nature and typically will be the amino acid sequence found in nature. In an aspect, the term "wild type" or "parental sequence" can indicate a starting or reference sequence prior to a manipulation of the invention.

In the polypeptide notation used herein, the left-hand direction is the amino terminal direction and the right-hand direction is the carboxy-terminal direction, in accordance with standard usage and convention. Similarly, unless specified otherwise, the

left-hand end of single-stranded polynucleotide sequences is the 5' end; the left-hand direction of double-stranded polynucleotide sequences is referred to as the 5' direction. The direction of 5' to 3' addition of nascent RNA transcripts is referred to as the transcription direction.

Detailed Description of the Invention

The present invention provides an in vitro method of making sequence variants from at least one heteroduplex polynucleotide wherein the heteroduplex has at least two non-complementary nucleotide base pairs, the method comprising: preparing at least one heteroduplex polynucleotide; combining said heteroduplex polynucleotide with an effective amount of an agent or agents with exonuclease activity, polymerase activity and strand cleavage activity; and allowing sufficient time for the percentage of complementarity to increase, wherein at least one or more variants are made.

Another aspect of the present invention is where the heteroduplex polynucleotides are circular, linear or a replicon.

Another aspect of the present invention is where the desired variants have different amounts of complementarity.

Another aspect of the present invention is where the exonuclease activity, polymerase activity, and strand cleavage activity is added sequentially, or concurrently.

Another aspect of the present invention provides the addition of ligase activity, provided by agents such as, T4 DNA ligase, E. coli DNA ligase, or Taq DNA ligase.

Another aspect of the present invention is where the strand cleavage activity is provided by an enzyme, such as, CEL I, T4 endonuclease VII, T7 endonuclease I, S1 nuclease, BAL-31 nuclease, FEN1, cleavase, pancreatic DNase I, SP nuclease, mung bean nuclease, and nuclease P1; a chemical agent, such as, potassium permanganate, tetraethylammonium acetate, sterically bulky photoactivatable DNA intercalators, [Rh(bpy)2(chrysi)]3+, osmium tetroxide with piperidine, and hydroxylamine with piperidine or a form of energy, such as, ionizing or kinetic radiation.

Another aspect of the present invention is where polymerase activity is provided by Pol beta.

Another aspect of the present invention is where both polymerase activity and 3' to 5' exonuclease activity is provided T4 DNA polymerase, T7 DNA polymerase, E. coli Pol 1, or Pfu DNA polymerase.

Another aspect of the present invention is where the agent with both polymerase activity and 5' to 3' exonuclease activity is E. coli Pol 1.

An embodiment of the present invention is where the effective amount of strand cleavage activity, and exonuclease activity/polymerase activity and ligase activity are provided by CEL I, T4 DNA polymerase, and T4 DNA ligase.

Another aspect of the present invention is where the effective amount of strand cleavage activity, and exonuclease activity/polymerase activity and ligase activity are provided by CEL I, T7 DNA polymerase, and T4 DNA ligase.

Another embodiment of the present invention provides an in vitro method of increasing diversity in a population of sequences, comprising, preparing at least one heteroduplex polynucleotide; combining the heteroduplex polynucleotide with an effective amount of an agent or agents with 3' to 5' exonuclease activity, polymerase activity and strand cleavage activity; and allowing sufficient time for the percentage of complementarity to increase, wherein diversity in the population is increased.

Another embodiment of the present invention provides a method of obtaining a polynucleotide encoding a desired functional property, comprising: preparing at least one heteroduplex polynucleotide; combining said

heteroduplex polynucleotide with an effective amount of an agent or agents with exonuclease activity, polymerase activity and strand cleavage activity; allowing sufficient time for the percentage of complementarity between strands of the heteroduplex polynucleotide to increase, wherein diversity in the population is increased; and screening or selecting a population of variants for the desired functional property.

Another embodiment of the present invention provides a method of obtaining a polynucleotide encoding a desired functional property, comprising: preparing at least one heteroduplex polynucleotide; combining said heteroduplex polynucleotide with an effective amount of an agent or agents with exonuclease activity, polymerase activity and strand cleavage activity; allowing sufficient time for the percentage of complementarity between strands of the heteroduplex polynucleotide to increase, wherein diversity in the population is increased; converting DNA to RNA; and screening or selecting a population of ribonucleic acid variants for the desired functional property.

Yet another embodiment of the present invention provides a method of obtaining a polypeptide having a desired functional property, comprising: preparing at least one heteroduplex polynucleotide; combining said heteroduplex polynucleotide with an effective amount of an agent or agents with exonuclease activity, polymerase activity and strand cleavage activity; allowing sufficient time for the percentage of complementarity

between strands of said heteroduplex polynucleotide to increase, converting said heteroduplex polynucleotide to RNA, and said RNA to a polypeptide; and screening or selecting a population of polypeptide variants for said desired functional property.

Still another embodiment of the present invention provides a method of obtaining a polynucleotide encoding a desired functional property, comprising: preparing at least one heteroduplex polynucleotide, where the heteroduplex is optionally, about 95%, 90%, 85%, 80%, or 75% identical, and about 1000KB, 10,000KB, or 100,000KB is size; combining said heteroduplex polynucleotide with an effective amount of an agent or agents with exonuclease activity, polymerase activity and strand cleavage activity; allowing sufficient time for the percentage of complementarity between strands of the heteroduplex polynucleotide to increase, screening or selecting for a population of variants having a desired functional property; denaturing said population of variants to obtain single strand polynucleotides; annealing said single strand polynucleotides to form at least one second heteroduplex polynucleotide; combining said second heteroduplex polynucleotide with an effective amount of an agent or agents with exonuclease activity, polymerase activity and strand cleavage activity; and allowing sufficient time for the percentage of complementarity between strands of the heteroduplex polynucleotide to increase.

The present invention is directed to a method for generating an improved polynucleotide sequence or a population of improved polynucleotide sequences, typically in the form of amplified and/or cloned polynucleotides, whereby the improved polynucleotide sequence(s) possess at least one desired phenotypic characteristic (*e.g.*, encodes a polypeptide, promotes transcription of linked polynucleotides, binds a protein, improves the function of a viral vector, and the like) which can be selected or screened for. Such desired polynucleotides can be used in a number of ways such as expression from a suitable plant, animal, fungal, yeast, or bacterial expression vector, integration to form a transgenic plant, animal or microorganism, expression of a ribozyme, and the like.

GRAMMR provides for a process where heteroduplexed DNA strands are created by annealing followed by resolution of mismatches in an *in vitro* reaction. This reaction begins with cleavage of one strand or the other at or near a mismatch followed by excision of mismatched bases from that strand and polymerization to fill in the resulting gap with nucleotides that are templated to the sequence of the other strand. The resulting nick can be sealed by ligation to rejoin the backbone. The sum of the separate independent mismatch resolution events that occur throughout the heteroduplex molecule will result in reassortment of sequence information throughout that molecule. The sequence information will reassort in a

variety of combinations to generate a complex library of sequence variants.

In one embodiment of GRAMMR, a library of mutants is generated by any method known in the art such as mutagenic PCR, chemical mutagenesis, *etc.* followed by screening or selection for mutants with a desired property. DNA is prepared from the chosen mutants. The DNAs of the mutants are mixed, denatured to single strands, and allowed to anneal. Partially complementary strands that hybridize will have non-base-paired nucleotides at the sites of the mismatches. Treatment with CEL I (Oleykowski *et al.*, 1998; Yang *et al.*, 2000), or a similar mismatch-directed activity, will cause nicking of one or the other polynucleotide strand 3' of each mismatch. (In addition, CEL I can nick 3' of an insertion/deletion resulting in reassortment of insertions/deletions.) The presence of a polymerase containing a 3'-to-5' exonuclease ("proofreading") activity (*e.g.*, T4 DNA Pol) will allow excision of the mismatch, and subsequent 5'-to-3' polymerase activity will fill in the gap using the other strand as a template. A polymerase that lacks 5′-3′ exonuclease activity and strand-displacement activity will fill in the gap and will cease to polymerize when it reaches the 5′ end of DNA located at the original CEL I cleavage site, thus re-synthesizing only short patches of sequence. Alternatively, the length of the synthesized patches can be modulated by spiking the reaction with a polymerase that contains a 5′-3′ exonuclease activity; this nick-translation activity can traverse a longer

region resulting in a longer patch of information transferred from the template strand. DNA ligase (*e.g.*, T4 DNA ligase) can then seal the nick by restoring the phosphate backbone of the repaired strand. This process can occur simultaneously at many sites and on either strand of a given heteroduplexed DNA molecule. The result is a randomization of sequence differences among input strands to give a population of sequence variants that is more diverse than the population of starting sequences. These output polynucleotides can be cloned directly into a suitable vector, or they can be amplified by PCR before cloning. Alternatively, the reaction can be carried out on heteroduplexed regions within the context of a double-stranded circular plasmid molecule or other suitable replicon that can be directly introduced into the appropriate host following the GRAMMR reaction. In another alternative, the output polynucleotides can be transcribed into RNA polynucleotides and used directly, for example, by inoculation of a plant viral vector onto a plant, such as in the instance of a viral vector transcription plasmid. The resulting clones are subjected to a selection or a screen for improvements in a desired property. The overall process can then be repeated one or more times with the selected clones in an attempt to obtain additional improvements.

If the output polynucleotides are cloned directly, there is the possibility of incompletely resolved molecules persisting that, upon replication in the cloning host, could lead to two different plasmids in the same cell. These plasmids could potentially give

rise to mixed-plasmid colonies. If it is desired to avoid such a possibility, the output polynucleotide molecules can be grown in the host to allow replication/resolution, the polynucleotides isolated and retransformed into new host cells.

In another embodiment, when sequence input from more than two parents per molecule is desired, the above procedure is performed in a cyclic manner before any cloning of output polynucleotides. After GRAMMR treatment, the double stranded polynucleotides are denatured, allowed to anneal, and the mismatch resolution process is repeated. After a desired number of such cycles, the output polynucleotides can be cloned directly, introduced into a suitable vector, or they can be amplified by PCR before cloning. The resulting clones are subjected to a selection or a screen for improvements in a desired property.

In another embodiment, a "molecular backcross" is performed to help eliminate the background of deleterious mutations from the desired mutations. A pool of desired mutants' DNA can be mixed with an appropriate ratio of wild-type DNA to perform the method. Clones can be selected for improvement, pooled, and crossed back to wild-type again until there is no further significant change.

The efficiency of the process is improved by various methods of enriching the starting population for heteroduplex molecules, thus reducing the number of

unaltered parental-type output molecules. The mismatched hybrids can be affinity purified using aptamers, dyes, or other agents that bind to mismatched DNA. A preferred embodiment is the use of MutS protein affinity matrix (Wagner *et al.*, Nucleic Acids Res. 23(19):3944-3948 (1995); Su *et al.*, Proc. Natl. Acad. Sci. (U.S.A.), 83:5057-5061(1986)) or mismatch-binding but non-cleaving mutants of phage T4 endonuclease VII (Golz and Kemper, Nucleic Acids Research, 1999; 27: e7).

In one embodiment, the procedure is modified so that the input polynucleotides consist of a single strand of each sequence variant. For example, single-stranded DNAs of opposite strandedness are produced from the different parent sequences by asymmetric PCR to generate partially complementary single-stranded molecules. Annealing of the strands with one-another to make heteroduplex is performed as described in Example 1. Alternatively, single-stranded DNAs can be generated by preferentially digesting one strand of each parental double-stranded DNA with Lambda exonuclease followed by annealing the remaining strands to one-another. In this embodiment, the annealing strands have no 100% complementary strand present with which to re-anneal. Hence, there is a lower background of unmodified polynucleotides, that is, "parental polynucleotides" among the output polynucleotides leading to a higher efficiency of reassorting sequence variations. This increased efficiency will be particularly valuable in situations where a screen rather than a selection is employed to test for the desired polynucleotides.

Another method for heteroduplex formation is to mix the double-stranded parent DNAs, denature to dissociate the strands, and allow the single-stranded DNAs to anneal to one-another to generate a population of heteroduplexes and parental homoduplexes. The heteroduplexes can then be selectively enriched by a heteroduplex capture method such as those described above using MutS or a non-cleaving T4 endonuclease VII mutant. Alternatively, the parental homoduplex molecules in the population may be cleaved by restriction enzymes that overlap with sites of mismatch such that they are not cleaved in the heteroduplex but are cleaved in the parental homoduplex molecules. Uncleaved heteroduplex DNA can then be isolated by size fractionation in an agarose gel as was performed to generate full-length plasmid on full-length plasmid heteroduplex DNA molecules as describe in Example 6. Circularization of those full-length heteroduplexed plasmid molecules was then brought about by incubation with DNA ligase.

In another embodiment, the parental, or input, double-stranded polynucleotides are modified by the addition of "clamp" sequences. One input polynucleotide or pool of polynucleotides is amplified by PCR with the addition of a unique sequence in the 5' primer. The other input polynucleotide or pool is amplified by PCR with the addition of a unique sequence in the 3' primer. The clamp sequences can be designed to contain a unique restriction enzyme site for the 5' end of the gene of interest and another for the 3' end such that, at the

step of cloning the products of the GRAMMR reassortment, only products with the 5' clamp from the first polynucleotide (or pool) and the 3' end from the second polynucleotide (or pool) will have appropriate ends for cloning. Alternatively, the products of GRAMMR reassortment can be PCR amplified using the unique sequences of the 5' and 3' clamps to achieve a similar result. Hence, there is a lower background of unmodified polynucleotides, that is, "parental polynucleotides" among the output polynucleotide clones leading to a higher efficiency of reassorting sequence variations. This increased efficiency will be particularly valuable in situations where a screen rather than a selection is employed to test for the desired polynucleotides. Optionally, oligonucleotide primers can be added to the GRAMMR reaction that are complementary to the clamp primer sequences such that either parent can serve as the top strand, thus permitting both reciprocal heteroduplexes to participate in the mismatch-resolution reaction.

Another method for generating cyclic heteroduplexed polynucleotides is performed where parental double-stranded DNAs have terminal clamp sequences as described above where the single-stranded clamp sequences extending from one end of the heteroduplex are complementary to single-stranded clamp sequences extending from the other end of the heteroduplex. These complementary, single-stranded clamps are allowed to anneal, thereby circularizing the heteroduplexed DNA molecule. Parental homoduplexes that

result from re-annealing of identical sequences have only one clamp sequence and therefore, no complementary single-stranded sequences at their termini with which circularization can occur. Additionally, a DNA polymerase and a DNA ligase can be used to fill-in any gaps in the circular molecules and to seal the nicks in the backbone, respectively, to result in the formation of a population of covalently-closed circular heteroduplex molecules. As the covalently-closed circular heteroduplex molecules will not dissociate into their component strands if subjected to further denaturating conditions, the process of denaturation, circularization, and ligation can be repeated to convert more of the linear double-stranded parental duplexes into closed into closed circular heteroduplexes.

In another embodiment, a region of a single-stranded circular phagemid DNA can be hybridized to a related, but non-identical linear DNA, which can then be extended with a polymerase such as T7 DNA polymerase or T4 DNA polymerase plus T4 gene 32 protein, then ligated at the resulting nick to obtain a circular, double-stranded molecule with heteroduplexed regions at the sites of differences between the DNAs. GRAMMR can then be carried out on this molecule to obtain a library of sequence-reassorted molecules.

Alternately, two single-stranded circular phagemid DNAs of opposite strand polarity relative to the plasmid backbone, and parent gene sequences that are the target of the reassortment are annealed to one and other.

A region of extensive mismatch will occur where the phage f1 origin sequences reside. Upon GRAMMR treatment, however, this region of extensive mismatch can revert to either parental type sequence restoring a function f1 origin. These double strained molecules will also contain mismatch regions at the sites of differences between the strands encoding the parent genes of interest. GRAMMR can then be carried out on this molecule to obtain a library of sequence re-assorted molecule.

As discussed in the preceding paragraphs, the starting DNA or input DNA can be of any number of forms. For example, input DNA can be full-length, single stranded and of opposite sense, as is taught in Example 1. Alternatively, the input DNA can also be a fragment of the full-length strand. The input DNAs can be double-stranded, either one or both, or modified, such as by, methylation, phosphorothiolate linkages, peptide-nucleic acid, substitution of RNA in one or both strands, or the like. Either strand of a duplex can be continuous along both strands, discontinuous but contiguous, discontinuous-with overlaps, or discontinuous with gaps.

GRAMMR can also be applied to DNA fragmentation and reassembly-based DNA shuffling schemes. For instance, in methods where gene fragments are taken through cycles of denaturation, annealing, and extension in the course of gene reassembly, GRAMMR can be employed as an intermediate step.

In one such embodiment, the DNA from a gene, or pool of mutants' genes is fragmented by enzymatic, mechanical or chemical means, and optionally a size range of said fragments is isolated by a means such as separation on an agarose gel. The starting polynucleotide, such as a wild-type, or a desired variant, or a pool thereof, is added to the fragments and the mixture is denatured and then allowed to anneal. The annealed polynucleotides are treated with a polymerase to fill in the single stranded gaps using the intact strand as a template. The resulting partially complementary double strands will have non-base-paired nucleotides at the sites of the mismatches. Treatment with CEL I (Oleykowski *et al.*, 1998; Yang *et al.*, 2000) will cause nicking of one or the other polynucleotide strand 3' of each mismatch. Addition of a polymerase containing a 3'-to-5' exonuclease that provides proofreading activity, such as, DNA Pol I, T4 DNA Pol I, will allow excision of the mismatch, and subsequent 5'-to-3' polymerase activity will fill in the gap using the other strand as a template. A DNA ligase, such as, T4 DNA Ligase, can then seal the nick by restoring the phosphate backbone of the repaired strand. The result is a randomization of sequence variation among input strands to give output strands with potentially improved properties. These output polynucleotides can be cloned directly into a suitable vector, or they can be amplified by PCR before cloning. The resulting clones are subjected to a selection or a screen for improvements in a desired property.

In one such embodiment, the DNA from a pool of mutants' genes is fragmented by enzymatic, mechanical or chemical means, or fragments are generated by limited extension of random oligonucleotides annealed to parental templates (United States Patent 5,965,408), and optionally a size range of said fragments is isolated by a means such as separation on an agarose gel. The mixture is denatured and then allowed to anneal. The annealed polynucleotides are optionally treated with a polymerase to fill in the single stranded gaps. The resulting partially complementary double-strand fragments will have non-base paired nucleotides at the sites of the mismatches. Treatment with CEL I (Oleykowski *et al.*, 1998; Yang *et al.*, 2000) will cause nicking of one or the other polynucleotide strand 3' of each mismatch. The activity of a polymerase containing a 3'-to-5' exonuclease ("proofreading") activity, such as T4 DNA Polymerase, will allow excision of the mismatch, and subsequent 5'-to-3' polymerase activity will fill in the gap using the other strand as a template. Optionally, DNA ligase, such as, T4 DNA Ligase, can then seal the nick by restoring the phosphate backbone of the repaired strand. The result is a randomization of sequence variation among input strands to give output strands with potentially improved properties. Subsequent rounds of denaturing, annealing, and GRAMMR treatment allows gene reassembly. PCR can be used to amplify the desired portion of the reassembled gene. These PCR output polynucleotides can be cloned into a suitable vector. The resulting clones are subjected to a selection or a screen for the desired functional property.

Another embodiment of the present invention provides starting with a continuous scaffold strand to which fragments of another gene or genes anneal. The flaps and gaps are trimmed and filled as is described in Coco, et al., Nature Biotech 19 (01)354; United States Patent 6,319,713, and GRAMMR is performed. In this process, GRAMMR would bring about further sequence reassortment by permitting transfer of sequence information between the template strand and the strand resulting from flap and gap trimming and ligation. This method provides the benefits of incorporating specific sequence patches into one continuous strand followed by GRAMMR of residues that mismatch with the scaffold. By annealing many fragments simultaneously to the same sequence or gene, many individual sites can be addressed simultaneously, thereby allowing reassortment of multiple sequences or genes at once. Unlike the method disclosed by Coco, et al., in the present embodiment, the scaffold is not degraded, rather the duplex can be directly cloned, or amplified by PCR prior to cloning. Exhaustive mismatch resolution will result in a perfectly duplexed DNA. Partial mismatch resolution will result in essentially two different reassorted products per duplex.

As can be appreciated from the present disclosure, GRAMMR can also be applied to a variety of methods that include the annealing of related DNAs as a step in their process. For example, many site-directed mutagenesis protocols call for the annealing of mutant-encoding DNA molecules to a circular DNA in single-

stranded form, either phagemid or denatured plasmid. These DNAs are then extended with a polymerase, followed by treatment with ligase to seal the nick, with further manipulation to remove the parental sequence, leaving the desired mutation or mutations incorporated into the parental genetic background. Though these protocols are generally used to incorporate specific mutations into a particular DNA sequence, it is feasible that the GRAMMR process can be applied to the heteroduplexed molecules generated in such a process to reassort sequence variations between the two strands, thereby resulting in a diverse set of progeny with reassorted genetic variation.

Another embodiment provides for a sequential round of reassortment on a particular region. For example, DNA fragments are annealed to a circular single-strand phagemid DNA, and GRAMMR is performed. The fragments can be treated in order to prevent them from being physically incorporated into the output material. For example, they can be terminated at the 3' end with di-deoxy residues making them non-extendible. Multiple rounds of reassortment can be performed, but only modified molecules from the original input single stranded DNA clone will be recovered. The consequence will be that the DNA fragments used in this reassortment will contribute only sequence information to the final product and will not be physically integrated into the final recoverable product.

In instances where it is desired to resolve only sites of significant mismatch, that is patches of more than about 1 to 3 mismatches, S1 nuclease can be used. S1 nuclease is an endonuclease specific for single-stranded nucleic acids. It can recognize and cleave limited regions of mismatched base pairs in DNA:DNA or DNA:RNA duplexes. A mismatch of at least about 4 consecutive base pairs is generally required for recognition and cleavage by S1 nuclease. Mismatch resolution will not occur if both strands are cleaved, so the DNA must be repaired after the first nick and before the counter-nick. Other nucleases may be preferable for specifically tuning cleavage specificity according to sequence, sequence context, or size of mismatch.

In addition, other means of addressing mismatched residues, such as chemical cleavage of mismatches may be used. Alternatively, one can choose to subject the strands of heteroduplexed DNA to random nicking with an activity such as that exhibited by DNaseI or an agent that cleaves only in duplexed regions. If nick formation occurs in a region of identity between the two genes, the DNA ligase present in the reaction will seal the nick with no net transfer of sequence information. However, if nick formation occurs near a site of mismatch, the mismatched bases can be removed by 3'-5' exonuclease and the gap filled in by polymerase followed by nick sealing by ligase. Alternatively, application of nick-translation through regions of heterogeneity can bring about sequence reassortment. These processes, though not directed exclusively by the

mismatch status of the DNA, will serve to transfer sequence information to the repaired strand, and thus result in a reassorted sequence.

GRAMMR can be used for protein, peptide, or aptamer display methods to obtain recombination between library members that have been selected. As fragmentation of the input DNAs is not required for GRAMMR, it may be possible to reassort sequence information between very small stretches of sequence. For instance, DNAs encoding small peptides or RNA aptamers that have been selected for a particular property such as target binding can be reassorted. For annealing to occur between the selected DNA molecules, some level of sequence homology should be shared between the molecules, such as at the 5' and 3' regions of the coding sequence, in regions of the randomized sequence segment that bear similarity because of similar binding activities, or through the biasing of codon wobble-base identity to a particular set of defaults.

Manipulation of the reaction temperature at which GRAMMR is conducted can be useful. For example, lower temperatures will help to stabilize heteroduplexes allowing GRAMMR to be performed on more highly mismatched substrates. Likewise, additives that affect base-pairing between strands, such as salts, PEG, formamide, etc, can be used to alter the stability of the heteroduplex in the GRAMMR, thereby affecting the outcome of the reaction.

In another embodiment, the mismatched double stranded polynucleotides are generated, treated with a DNA glycosylase to form an apurinic or apyrimidinic site, (that is an "AP site") an AP endonuclease activity to cleave the phosphodiester bond, deoxyribulose phosphodiesterase to remove the deoxyribose-phosphate molecules, DNA polymerase ß or other DNA polymerase to add a single nucleotide to the 3' end of the DNA strand at the gap, and DNA ligase to seal the gap. The result is a reassortment of sequence variations between input strands to give output strands with potentially improved properties. These output polynucleotides can be cloned directly into a suitable vector, or they can be amplified by PCR before cloning. The resulting clones are subjected to a selection or a screen for improvements in a desired property.

Another embodiment provides for zonal mutagenesis by GRAMMR, that is, random or semi-random mutations at, and in the immediate vicinity of, mismatched residues using nucleotide analogues that have multiple base-pairing potential. This provides for concentration of essentially random mutagenesis at a particular point of interest, and adds another benefit to the present invention. Similar genes with slightly different functions, for example, plant R-genes, enzymes, or the like, will exhibit moderate sequence differences between them in regions that will be important for their own particular activities. Genes that express these activities, such as different substrates, binding partners, regulatory sites, or the like, should have

heterogeneity in the regions that govern these functions. Since it is known that the specificity of such functions is associated with these amino acids and their neighbors, GRAMMR mutagenesis might serve to both reassort sequence variation among genes and also direct random mutagenesis to these regions to drive them further and faster evolutionarily, while not disturbing other sequences, such as structural framework, invariant residues, and other such important sites, that are potentially less tolerant to randomization.

Different enzymes with distinct functions will not differ just in the operative regions, such as active sites, regulatory sites, and the like. They are likely to have other differences from one another that arise through genetic drift. Further randomization in the locales of such changes might therefore be considered neutral, minimally important, or deleterious to the outcome of a mutagenesis experiment. In order to direct the random mutagenesis away from such inconsequential sites, and toward sites that might present a better result for random mutagenesis, such as the active site of an enzyme, the codon usage bias of the genes could be manipulated to decrease or increase the overall level of nucleotide complementarity in those regions. If regions of greater complementarity are less susceptible to GRAMMR than regions of lesser complementarity, then the degree of GRAMMER-directed zonal random mutagenesis at a given site can be modulated.

In another embodiment, after heteroduplex molecules are formed, an enzyme with a 3' to 5' exonuclease activity is added such that one strand of each end of the heteroduplex is digested back. At a point at which, on average, a desired amount of 3' to 5' digestion has occurred, dNTPs are added to allow the 5' to 3' polymerase activity from the same or an additional enzyme to restore the duplex using the opposite strand as a template. Thus mismatches in the digested regions are resolved to complementarity. Optionally, the resultant duplexes are purified, denatured and then allowed to anneal. The process of digestion, then polymerization is repeated resulting in new chimeric sequences. Additional cycles of the process can be performed as desired. Output duplex molecules are cloned and tested for the desired functional property. This process requires no fragmentation and reassembly. In addition, this process requires no endonucleolytic cleavages.

In another embodiment, after the heteroduplex molecules are formed, an enzyme with a 5' to 3' exonuclease activity, such as, T7 Gene6 Exonuclease as disclosed in Enger, MJ and Richardson, CC , J Biol Chem 258(83)11197), is added such that one strand of each end of the heteroduplex is digested. At a point at which, on average, a desired amount of 5' to 3' digestion has occurred, the reaction is stopped and the exonuclease inactivated. Oligonucleotide primers complementary to the 5' and 3' ends of the target polynucleotides are added and annealed. A DNA polymerase, such as, T4 DNA Polymerase, a DNA ligase and dNTPs are added to allow the

5' to 3' polymerase activity to extend the primers and restore the duplex using the opposite strand as a template, with ligase sealing the nick. Thus mismatches in the digested regions are resolved to complementarity. Optionally, the resultant duplexes are purified, denatured and then allowed to anneal. The process of digestion then polymerization is repeated resulting in new chimeric sequences. Additional cycles of the process can be performed as desired. Output duplex molecules are cloned and tested for the desired functional property. This process requires no fragmentation and reassembly. In addition, this process requires no endonucleolytic cleavages.

In the current invention the random reassortment occurs in an *in vitro* DNA mismatch-resolution reaction. This method does not require any steps of "gene reassembly" that serve as the foundation for the earlier mutation reassortment ("shuffling") methods. Instead, it is based upon the ability of a reconstituted or artificial DNA mismatch resolving system to transmit sequence variations from one or more strands of DNA into another DNA strand by hybridization and mismatch resolution *in vitro*.

In general, standard techniques of recombinant DNA technology are described in various publications, *e.g.*, (Ausubel, 1987; Ausubel, 1999; Sambrook *et al.*, 1989), each of which is incorporated herein in their entirety by reference. Polynucleotide modifying enzymes were used according to the manufacturers recommendations.

If desired, PCR amplimers for amplifying a predetermined DNA sequence may be chosen at the discretion of the practitioner.

It is noted that each of the activities taught in the present invention that are involved in the GRAMMR reaction can be interchanged with a functional equivalent agent with similar activity, and that such changes are within the scope of the present invention. For instance, as was indicated in Example 2, Taq DNA ligase could substitute for T4 DNA ligase. Other ligases can be substituted as well, such as *E. coli* DNA ligase. Likewise, as shown in Examples 2 and 8, respectively, Pfu polymerase and T7 DNA polymerase can be substituted for T4 DNA polymerase. Other enzymes with appropriate exonuclease activity with or without associated polymerase can function in place of any of these enzymes for the exonuclease activity needed for the GRAMMR reaction. In a similar way, any polymerase with functionally equivalent activity to those demonstrated to work for GRAMMR can be used for substitution. These include *E. coli* Pol 1, the Klenow fragment of *E. coli* Pol 1, polymerase beta, among many others.

Strand cleavage may be brought about in a number of ways. In addition to CEL I, a number of functionally equivalent, and potentially homologous activities found in extracts from a variety of plant species (Oleykowski, Nucleic Acids Res 1998;26:4597-602) may be used. Other mismatch-directed endonucleases such as T4 endonuclease VII, T7 endonuclease I, and SP

nuclease (Oleykowski, Biochemistry 1999; 38: 2200-5) may be used. Other nucleases which attack single stranded DNA can be used, such as S1 nuclease, FEN1, cleavase, mung bean nuclease, and nuclease P1. Enzymes that make random cleavage events in DNA, such as pancreatic DNase I may also be substituted for the strand cleaving activity in GRAMMR. A number of methods for bringing about strand cleavage through other means are also envisioned. These include potassium permanganate used with tetraethylammonium acetate, the use of sterically bulky photoactivatable DNA intercalators such as [Rh(bpy)2(chrysi)]3+, osmium tetroxide with piperidine alkaloid, and hydroxylamine with piperidine alkaloid, as well as the use of radiation energy to bring about strand breakage.

CEL I is a mismatch endonuclease isolated from celery. The use of CEL I in a diagnostic method for the detection of mutations in targeted polynucleotide sequences, in particular, those associated with cancer, is disclosed in U.S. Patent No. 5,869,245. Methods of isolating and preparing CEL I are also disclosed in this patent. However, there is no disclosure in this patent relating to the use of CEL I in DNA sequence reassortment.

The nucleic acid molecules that encode CEL I are disclosed in PCT Application Publication No. WO 01/62974 A1. As with U.S. Patent No. 5,869,245, the use of CEL I in a diagnostic method for the detection of mutations in targeted polynucleotide sequences associated

with cancer is disclosed. Also similarly, there is no disclosure relating to the use of CEL I in DNA reassortment.

The reactivity of Endonuclease VII of phage T4 with DNA-loops of eight, four, or one nucleotide, or any of 8 possible base mismatches *in vitro* is disclosed in "Endonuclease VII of Phage T4 Triggers Mismatch Correction *in Vitro*" Solaro, et al., J Mol Biol 230(93)868. The publication reports a mechanism where Endonuclease VII introduces double stranded breaks by creating nicks and counternicks within six nucleotides 3' of the mispairing. The publication discloses that a time delay between the occurrence of the first nick and the counternick was sufficient to allow the 3'-5' exonuclease activity of gp43 to remove the mispairing and its polymerase activity to fill in the gap before the occurrence of the counternick. Nucleotides are erased from the first nick, which is located 3' of the mismatch on either strand and stops 5' of the mismatch at the first stable base-pair. The polymerase activity proceeds in the 5' to 3' direction towards the initial nick, which is sealed by DNA ligase. As a result, very short repair tracks of 3 to 4 nucleotides extend across the site of the former mismatch. The publication concludes with a discussion regarding the various activities Endonuclease VII may have within phage T4. However, the publication does not disclose any practical utility for Endonuclease VII outside of phage T4, and there is no disclosure regarding its applicability in DNA reassortment.

A method for creating libraries of chimeric DNA sequences *in vivo* in *Escherichia coli* is disclosed in *Nucleic Acids Research*, 1999, Vol 27, No. 18, e18, Volkov, A. A., Shao, Z., and Arnold, F.H. The method uses a heteroduplex formed *in vitro* to transform *E. coli* where repair of regions of non-identity in the heteroduplex creates a library of new, recombined sequences composed of elements of each parent. Although the publication discloses the use of this method as a convenient addition to existing DNA recombination methods, that is, DNA shuffling, the disclosed method is limited to the *in vivo* environment of *E. coli*. The publication states that there is more than one mechanism available for mismatch repair in *E. coli*, and that the 'long patch' repair mechanism, which utilizes the MutS/L/H enzyme system, was probably responsible for the heteroduplex repair.

Cited References

1. Arkin, A.P. and Youvan, D.C. (1992) An algorithm for protein engineering: simulations of recursive ensemble mutagenesis. *Proc Natl Acad Sci U S A*, **89**, 7811-7815.

2. Ausubel, F.M. (1987) *Current protocols in molecular biology*. Published by Greene Pub. Associates and Wiley-Interscience : J. Wiley, New York.

3. Ausubel, F.M. (1999) *Short protocols in molecular biology : a compendium of methods from Current protocols in molecular biology*. Wiley, New York.

4. Barnes, W.M. (1994) PCR amplification of up to 35-kb DNA with high fidelity and high yield from lambda bacteriophage templates. *Proc Natl Acad Sci U S A*, **91**, 2216-2220.

5. Bartel, D.P. and Szostak, J.W. (1993) Isolation of new ribozymes from a large pool of random sequences. *Science*, **261**, 1411-1418.

6. Cadwell, R.C. and Joyce, G.F. (1992) Randomization of genes by PCR mutagenesis. *PCR Methods Appl*, **2**, 28-33.

7. Calogero, S., Bianchi, M.E. and Galizzi, A. (1992) In vivo recombination and the production of hybrid genes. *FEMS Microbiol Lett*, **76**, 41-44.

8. Caren, R., Morkeberg, R. and Khosla, C. (1994) Efficient sampling of protein sequence space for multiple mutants. *Biotechnology (N Y)*, **12**, 517-520.

9. Delagrave, S., Goldman, E.R. and Youvan, D.C. (1993) Recursive ensemble mutagenesis. *Protein Eng*, **6**, 327-331.

10. Delagrave, S. and Youvan, D.C. (1993) Searching sequence space to engineer proteins: exponential ensemble mutagenesis. *Biotechnology (N Y)*, **11**, 1548-1552.

11. Goldman, E.R. and Youvan, D.C. (1992) An algorithmically optimized combinatorial library screened by digital imaging spectroscopy. *Biotechnology (N Y)*, **10**, 1557-1561.

12. Gram, H., Marconi, L.A., Barbas, C.F.d., Collet, T.A., Lerner, R.A. and Kang, A.S. (1992) In vitro selection and affinity maturation of antibodies from a naive combinatorial immunoglobulin library. *Proc Natl Acad Sci U S A*, **89**, 3576-3580.

13. Hayashi, N., Welschof, M., Zewe, M., Braunagel, M., Dubel, S., Breitling, F. and Little, M. (1994) Simultaneous mutagenesis of antibody CDR regions by overlap extension and PCR. *Biotechniques*, **17**, 310, 312, 314-315.

14. Hermes, J.D., Blacklow, S.C. and Knowles, J.R. (1990) Searching sequence space by definably random mutagenesis: improving the catalytic potency of an enzyme. *Proc Natl Acad Sci U S A*, **87**, 696-700.

15. Holland, J.H. (1992) *Adaptation in natural and artificial systems : an introductory analysis with applications to biology, control, and artificial intelligence*. MIT Press, Cambridge, Mass.

16. Ji, G. and Silver, S. (1992) Regulation and expression of the arsenic resistance operon from Staphylococcus aureus plasmid pI258. *J Bacteriol*, **174**, 3684-3694.

17. Kauffman, S.A. (1993) *The origins of order : self-organization and selection in evolution*. Oxford University Press, New York.

18. Marton, A., Delbecchi, L. and Bourgaux, P. (1991) DNA nicking favors PCR recombination. *Nucleic Acids Res*, **19**, 2423-2426.

19. Meyerhans, A., Vartanian, J.P. and Wain-Hobson, S. (1990) DNA recombination during PCR. *Nucleic Acids Res*, **18**, 1687-1691.

20. Nissim, A., Hoogenboom, H.R., Tomlinson, I.M., Flynn, G., Midgley, C., Lane, D. and Winter, G. (1994) Antibody fragments from a 'single pot' phage display library as immunochemical reagents. *Embo J*, **13**, 692-698.

21. Oleykowski, C.A., Bronson Mullins, C.R., Godwin, A.K. and Yeung, A.T. (1998) Mutation detection using a novel plant endonuclease. *Nucleic Acids Res*, **26**, 4597-4602.

22. Oliphant, A.R., Nussbaum, A.L. and Struhl, K. (1986) Cloning of random-sequence oligodeoxynucleotides. *Gene*, **44**, 177-183.

23. Sambrook, J., Maniatis, T. and Fritsch, E.F. (1989) *Molecular cloning : a laboratory manual*. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.

24. Stemmer, W.P. (1994a) DNA shuffling by random fragmentation and reassembly: in vitro recombination for molecular evolution. *Proc Natl Acad Sci U S A*, **91**, 10747-10751.

25. Stemmer, W.P. (1994b) Rapid evolution of a protein in vitro by DNA shuffling. *Nature*, **370**, 389-391.

26. Stemmer, W.P., Morris, S.K. and Wilson, B.S. (1993) Selection of an active single chain Fv antibody from a protein linker library prepared by enzymatic inverse PCR. *Biotechniques*, **14**, 256-265.

27. Winter, G., Griffiths, A.D., Hawkins, R.E. and Hoogenboom, H.R. (1994) Making antibodies by phage display technology. *Annu Rev Immunol*, **12**, 433-455.

28. Yang, B., Wen, X., Kodali, N.S., Oleykowski, C.A., Miller, C.G., Kulinski, J., Besack, D., Yeung, J.A., Kowalski, D. and Yeung, A.T. (2000) Purification, cloning, and characterization of the CEL I nuclease. *Biochemistry*, **39**, 3533-3541.

The following non-limiting examples are provided to illustrate the present invention.

EXAMPLE 1

Cleavage of Mismatched DNA Substrate by CEL I

This example teaches the preparation of CEL I enzyme and its use in the cleavage of mismatched DNA substrate.

CEL I enzyme was prepared from celery stalks using the homogenization, ammonium sulfate, and Concanavalin A-Sepharose protocol described by Yang et al. (Biochemistry, 39:3533-3541 (2000), incorporated herein by reference. A 1.5 kg sample of chilled celery stalks was homogenized with a juice extractor. One liter of juice was collected, adjusted to 100 mM Tris-HCL, pH 7.7 with 100 micromolar phenylmethylsulfonyl fluoride (PMSF), and filtered through two layers of miracloth. Solid $(NH_4)_2SO_4$ was slowly added to 25% saturation while stirring on ice. After 30 minutes, the suspension was centrifuged at 27,000g for 1.5 hours at 4°C. The supernatants were collected and adjusted with solid $(NH_4)_2SO_4$ to 80% saturation while stirring on ice followed by centrifugation at 27,000g for 2 hours. The pellets were re-suspended in buffer B (0.1 M Tris-HCL, pH 7.7, 0.5 M KCl, 100 micromolar PMSF) and dialyzed against the same buffer.

Conconavalin A (ConA) Sepharose affinity chromatography was performed by first incubating the dialyzed sample with 2 ml of ConA resin overnight with

gentle agitation. The ConA resin was then packed into a 0.5 cm diameter column and washed with several column volumes of buffer B. Elution was performed using 0.3 M alpha-methyl-mannoside in buffer B. Fractions were collected in 1 ml aliquots. Fractions were assayed for mismatch cleavage activity on a radiolabeled mismatch substrate by incubating 0.1 microliter of each fraction with the mismatched probe in buffer D (20 mM Tris-HCL, pH 7.4, 25 mM KCL, 10 mM $MgCl_2$) for 30 minutes at 45°C as described by Oleykowski et al. (Nucleic Acids Research 26: 4597-4602 (1998), incorporated herein by reference. Reaction products were visualized by separation on 10% TBE-PAGE gels containing 7% urea (Invitrogen), followed by autoradiography. Aliquots of the CEL I fractions having mismatch cleavage activity were stored frozen at -20°C. A series of five-fold dilutions of CEL I fraction #5 were then analyzed for mismatch cleavage of radiolabeled mismatch substrate. Reactions were performed either in buffer D, New England BioLabs (NEB) T4 DNA ligase buffer (50 mM Tris-HCL, pH 7.5, 10 mM $MgCl_2$, 10 mM dithiothreitol (DTT), 1 mM ATP, 25 microgram/ml BSA), or Gibco/BRL T4 DNA ligase buffer (50 mM Tris-HCL, pH 7.6, 10mM $MgCl_2$, 1 mM DTT, 1 mM ATP, 5%(w/v) polyethylene glycol-8000). Reaction products were visualized as above. Cleavage activity in buffer D and in NEB T4 DNA ligase buffer were found to be roughly equivalent, whereas cleavage in the PEG-containing Gibco/BRL ligase buffer was enhanced by five to ten-fold compared to the other buffers.

Additional analysis of CEL I activity was carried out using defined heteroduplex DNAs from two different Green Fluorescent Protein (GFP) genes as substrate. This GFP heteroduplex substrate was prepared by annealing single stranded DNAs corresponding to cycle 3 GFP on the sense strand and wild-type GFP on the antisense strand. The single-stranded DNAs had been synthesized by asymmetric PCR and isolated by agarose gel electrophoresis. After annealing by heating to 90°C and cooling in the presence of 1X NEB restriction enzyme buffer 2 (10 mM Tris-HCL, pH 7.9, 10 mM $MgCl_2$, 50 mM NaCl, 1 mM dithiothreitol), the heteroduplex DNA was isolated by agarose gel electrophoresis followed by excision of the heterduplex band and extraction using Qiaquick DNA spin columns. A total of twenty eight mismatches, one or two nucleotides in length, occur throughout the length of the heteroduplex molecule. The distribution of the mismatches ranges from small clusters of several mismatches separated by one or two nucleotides to mismatches separated by more than thirty base pairs on either side.

A series of three-fold dilutions of CEL I in 1X NEB T4 DNA ligase buffer were prepared and one microliter aliquots of each were incubated in two separate series of 10 microliter reactions, each containing as substrate either 0.5 microgram of a supercoiled plasmid preparation or one hundred nanograms of the cycle3/wild-type GFP heteroduplex. All reactions took place in 1X NEB T4 DNA ligase buffer. Reactions were incubated at 45°C for 30

minutes and run on 1.5% TBE-agarose gel in the presence of ethidium bromide.

Treatment of the supercoiled plasmid preparation with increasing amounts of CEL I resulted in the conversion of supercoiled DNA to nicked circular, then linear molecules, and then to smaller fragments of DNA of random size. Treatment of the mismatched GFP substrate with the CEL I preparation resulted in the digestion of the full-length heteroduplex into laddered DNA bands which are likely to represent cleavage on opposite DNA strands in the vicinity of clusters of mismatches. Further digestion resulted in the conversion of the mismatched GFP substrate to smaller DNAs that may represent a limit digest of the heteroduplex DNA by the CEL I preparation.

EXAMPLE 2

Conservation of Full Length GFP Gene with Mismatch Resolution Cocktails

This example teaches various mismatch resolution cocktails that conserve the full length GFP Gene.

Mismatched GFP substrate was treated with various concentrations of CEL I in the presence of cocktails of enzymes that together constitute a synthetic mismatch resolution system. The enzymes used were CEL I, T4 DNA polymerase, Taq DNA polymerase and T4 DNA ligase.

CEL I activity should nick the heteroduplex 3' of mismatched bases. T4 DNA polymerase contains 3'-5' exonuclease for excision of the mismatched base from the nicked heteroduplex. T4 DNA polymerase and Taq DNA polymerase contain DNA polymerase capable of filling the gap. T4 DNA ligase seals the nick in the repaired molecule. Taq DNA polymerase also has 5' flap-ase activity.

Matrix experiments were performed to identify the reaction conditions that would serve to resolve mismatches in the GFP heteroduplex substrate. In one experiment, cycle 3/wild-type GFP heteroduplex was incubated in a matrix format with serial dilutions of CEL I fraction number five (described above) at eight different concentrations. Each reaction contained 100 nanograms of heteroduplex substrate and 0.2 microliters of T4 DNA ligase (Gibco BRL) in 1X NEBT4 DNA ligase buffer and dNTPs at 250 micromolar each, in a reaction volume of 10 microliters. In all, the matrix contained 96 individual reactions. One full set of reactions was incubated at room temperature for 30 minutes while another full set was incubated at 37°C for 30 minutes.

After incubation, PCR was used to amplify the GFP gene from each reaction. Aliquots from each PCR were then digested with *Hin*dIII and *Hpa*I and electrophoresed on 3% agarose gels with ethidium bromide. Only cycle 3 GFP has a *Hin*dIII site and only wild-type encodes a *Hpa*I site.

If DNA mismatch resolution occurred at either the *Hin*dIII or *Hpa*I mismatched sites, then a proportion of the PCR product would be expected to contain both sites, yielding a novel band. The band was observed in all samples, including the negative control samples that had neither CEL I, nor T4 DNA polymerase, nor Taq DNA polymerase. The results suggested that a basal level of background recombination may have occurred at some point in the experiment other than in the GRAMMR reaction; possibly in the PCR step. PCR-mediated recombination is known to occur at some frequency between related sequences during amplification [reference Paabo, et al., DNA damage promotes jumping between templates during enzymatic amplification. J Biol Chem 265(90)4718-4721].

In another experiment, 200 nanograms of cycle 3/wild-type GFP heteroduplex was treated with CEL I and T4 DNA polymerase in various concentrations along with 2.5 units of Taq DNA polymerase in the presence or absence of T4 DNA ligase (0.2 units;Gibco BRL). Each reaction contained 1X NEB T4 DNA ligase buffer with 0.05 mM each dNTP in a final volume of 20 microliters. Reactions were incubated for 30 minutes at 37°C and 10 microliters were run on a 2% TBE-agarose gel in the presence of ethidium bromide. Results showed that in the presence of DNA ligase, but in the absence of T4 DNA polymerase, increasing amounts of CEL I caused greater degradation of the heteroduplexed DNA, but that this effect could be counteracted by increasing the amount of T4 DNA polymerase in the reaction. These results indicated that the various components of the complete

reaction could act together to conserve the integrity of the full-length gene through DNA mismatch resolution.

Another matrix experiment was conducted to expand on these results and to identify additional conditions for DNA mismatch resolution for this synthetic system. 60 nanograms of cycle3/wild-type GFP heteroduplex were treated with CEL I and T4 DNA polymerase at various concentrations in the presence of 2.5 units of Taq DNA polymerase and 0.2 units of T4 DNA ligase in 1 X NEB T4 DNA ligase buffer containing 0.5 mM of each dNTP in a reaction volume of 10 microliters. Each set of reactions was incubated for 1 hour at either 20°C, 30°C, 37°C, or at 45°C. All reactions were then run on a 1.5% TBE-agarose gels in the presence of ethidium bromide. The results showed that the GFP heteroduplex was cleaved into discrete fragments by the CEL I preparation alone. The success of DNA mismatch resolution was initially gauged by the degree to which the apparent full-length integrity of the GFP sequence was maintained by the other components of the mismatch resolution system in the presence of CEL I. Conditions of enzyme concentration and temperature were identified that conserved a high proportion of the DNA as full-length molecules in this assay. Namely, one microliter of the CEL I fraction five preparation (described in Example 1) with one microliter (1 unit) of the T4 DNA polymerase in the presence of the other reaction components which were held constant in the experiment. It was found that as the reaction temperature increased, the degradative activity of CEL I increased accordingly.

Furthermore, it was shown that the other components of the repair reaction acted to conserve the integrity of the full-length DNA at 20°C, 30°C, and 37°C, but was remarkably less efficient at conserving the full-length DNA at 45°C. From these results, we concluded that under these experimental conditions, incubation at 45°C was not optimal for the process of GRAMMR, and that incubation at 20°C, 30°C, and 37°C were permissible.

Another experiment was performed in which alternative enzymes were used for the DNA mismatch resolution reaction. Instead of T4 DNA ligase, Taq DNA ligase was used. Pfu DNA polymerase (Stratagene) was employed in a parallel comparison to a set of reactions that contained T4 DNA polymerase as the 3' exonuclease/polymerase. Reactions were carried out in Taq DNA ligase buffer containing 8 units of Taq DNA ligase (NEB), 2.5 units Taq DNA polymerase, 0.5 mM of each dNTP, various dilutions of CEL I, and either T4 DNA polymerase or Pfu DNA polymerase). Reactions were run on a 1.5% TBE-agarose gels in the presence of ethidium bromide. It was found that in the presence of the Pfu DNA polymerase, Taq DNA polymerase, and Taq DNA ligase, the full-length integrity of the CEL I-treated substrate DNA was enhanced compared to DNA incubated with CEL I alone. This result shows that enzymes with functionally equivalent activities can be successfully substituted into the GRAMMR reaction.

Example 3

Restoration of Restriction Sites to GFP Heteroduplex DNA after DNA Mismatch Resolution (GRAMMR)

This experiment teaches the operability of genetic reassortment by DNA mismatch resolution (GRAMMR) by demonstrating the restoration of restriction sites.

The full-length products of a twenty-fold scale-up of the GRAMMR reaction, performed at 37°C for one hour, using the optimal conditions found above (the 1x reaction contained sixty nanograms of heteroduplex DNA, one microliter of CEL I fraction five (described in Example 1), one unit T4 DNA polymerase in the presence of 2.5 units of Taq DNA polymerase and 0.2 units of T4 DNA ligase in 1 X NEB T4 DNA ligase buffer containing 0.5 mM of each dNTP in a reaction volume of 10 microliters) were gel-isolated and subjected to restriction analysis by endonucleases whose recognition sites overlap with mismatches in the GFP heteroduplex, thereby rendering those sites in the DNA resistant to restriction enzyme cleavage. The enzymes used were *Bam*HI, *Hin*dIII, *Hpa*I, and *Xho*I. Negative controls consisted of untreated GFP heteroduplex. Positive controls consisted of Cycle 3 or wild type GFP sequences, individually. All controls were digested with the same enzymes as the product of the DNA mismatch resolution reaction. All samples were run on a 2% TBE-agarose gel in the presence of ethidium bromide.

After treatment with the mismatch resolution cocktail, a proportion of the DNA gained sensitivity to *Bam*HI and *Xho*I restriction endonucleases, indicating that DNA mismatch resolution had occurred. The *Hpa*I-cut samples could not be interpreted since a low level of cleavage occurred in the negative control. The *Hin*dIII, *Bam*HI and *Xho*I sites displayed different degrees of cleavage in the GRAMMR-treated samples. Restoration of the *Xho*I site was more extensive than that of the *Bam*HI site, which was in turn, more extensive than restoration at *Hin*dIII site.

The extent to which cleavage occurs is indicative of the extent to which mismatches in the DNA have been resolved at that site. Differences in mismatch resolution efficiency may relate to the nature or density of mismatches present at those sites. For example, the *Xho*I site spans a three-mismatch cluster, whereas the *Bam*HI site spans two mismatches and the *Hin*dIII site spans a single mismatch.

Example 4

GRAMMR-Reassorted GFP Genes

This example demonstrates that GRAMMR can reassort sequence variation between two gene sequences in a heteroduplex and that there are no significant differences in GRAMMR products that were directly cloned, or PCR amplified prior to cloning.

The GRAMMR-treated DNA molecules of Example 3 were subsequently either directly cloned by ligation into pCR-Blunt II-TOPO (Invitrogen), or amplified by PCR and ligated into pCR-Blunt II-TOPO according to the manufacturer's instructions, followed by transformation into *E. coli*. After picking individual colonies and growing in liquid culture, DNA was prepared and the sequences of the GFP inserts were determined. As negative controls, the untreated GFP heteroduplex substrate was either directly cloned or PCR amplified prior to cloning into the plasmid.

In GRAMMR, reassortment of sequence information results from a process of information transfer from one strand to the other. These sites of information transfer are analogous to crossover events that occur in recombination-based DNA shuffling methods. For the purposes of relating the results of these reassortment experiments, however, the GRAMMR output sequences are described in terms of crossovers. Sequences of twenty full-length GFP clones that were derived from the GRAMMR-treated GFP genes were analyzed. Four of these clones were derived from DNA that had been directly cloned into pZeroBlunt [ref] following GRAMMR treatment (no PCR amplification). The other sixteen sequences were cloned after PCR amplification. Analysis of these full-length GFP sequences revealed that all twenty sequences had undergone sequence reassortment having between one and ten crossovers per gene. A total of 99 crossovers were found in this set of genes, giving an average of about 5 crossovers per gene. With the distance between the first

and last mismatches of about 590 nucleotides, an overall frequency of roughly one crossover per 120 base-pairs was calculated. Within this set of twenty clones, a total of seven point mutations had occurred within the sequences situated between the PCR primer sequences, yielding a mutation frequency of roughly 0.05%.

Thirty-five clones that had not been subjected to GRAMMR treatment were sequenced. Of these controls, fourteen were derived from direct cloning and twenty-one were obtained after PCR amplification using the GFP heteroduplex as template. Of these thirty-five non-GRAMMR treated control clones, eight were recombinants, ranging from one to three crossovers, with most being single crossover events. A total of twenty-five point mutations had occurred within the sequences situated between the PCR primers, yielding a mutation frequency of roughly 0.1%.

No significant differences were observed between the GRAMMR-treated products that were either directly cloned or PCR amplified. Notably, though, in the non-GRAMMR-treated controls, the frequency of recombinants was higher in the PCR amplified DNAs than in the directly cloned DNAs. This higher frequency is consistent with results obtained by others in which a certain level of recombination was found to be caused by "jumping PCR." [Paabo, *et al.*, DNA damage promotes jumping between templates during enzymatic amplification. J Biol Chem 265(90)4718-4721].

Example 5

Heteroduplex Substrate Preparation for Plasmid-on-Plasmid Genetic Reassortment By DNA Mismatch Resolution (POP GRAMMR) of GFP Plasmids

This example teaches that heteroduplex substrate for Genetic Reassortment by DNA Mismatch Resolution can be in the form of intact circular plasmids. Cycle 3-GFP and wild-type GFP heteroduplex molecules were prepared plasmid-on-plasmid (POP) format. In this format, the GFP sequences were reassorted within the context of a circular double-stranded plasmid vector backbone. This made possible the recovery of the reassorted product by direct transformation of *E. coli* using an aliquot of the GRAMMR reaction. Consequently, neither PCR amplification nor other additional manipulation of the GRAMMR-treated DNA was necessary to obtain reassorted clones.

Mismatched DNA substrate for POP-GRAMMR reactions was generated containing wild-type GFP (SEQ ID NO:01) and Cycle 3 GFP (SEQ ID NO:02), resulting in the two pBluescript-based plasmids, pBSWTGFP (SEQ ID NO:03) and pBSC3GFP (SEQ ID NO:04), respectively. The GFPs were inserted between the *Kpn*I and *Eco*RI sites of the pBluescript polylinker so that the only sequence differences between the two plasmids occurred at sites where the wild-type and Cycle 3 GFPs differ from one-another. Both plasmids were linearized by digestion of

the plasmid backbone with *Sap*I, cleaned up using a DNA spin-column, mixed, amended to 1X PCR buffer (Barnes, 1994; *PNAS*, **91**, 2216-2220), heated in a boiling water bath for three minutes, and slow-cooled to room temperature to anneal the denatured DNA strands. Denaturing and annealing these DNAs led to a mixture of duplexes, the re-formation of parental duplexes, and the formation of heteroduplexes from the annealing of strands from each of the two input plasmids. Parental duplexes were deemed undesirable for GRAMMR and were removed by digestion with restriction enzymes that cut in one or the other parental duplex but not in the heteroduplexed molecules. *Pml*I and *Xho*I were chosen for this operation since *Pml*I cuts only in the wild-type GFP sequence and *Xho*I cuts only Cycle 3 GFP. After treatment with these enzymes, the products were resolved on an agarose gel. The full-length, uncut heteroduplex molecules were resolved from the *Pml*I- and *Xho*I-cut parental homoduplexes in an agarose gel and purified by excision of the band and purification with a DNA spin column.

The resulting population of heteroduplexed molecules was treated with DNA ligase to convert the linear DNA into circular, double-stranded DNA heteroduplexes. After confirmation by agarose gel-shift analysis, the circular double-stranded GFP heteroduplexed plasmid was used as substrate for GRAMMR reactions. Examples of the resulting clones are included as SEQ ID NO:05, SEQ ID NO:06, SEQ ID NO:07, and SEQ ID NO:08.

Example 6

Exemplary Reaction Parameters
for Genetic Reassortment by DNA Mismatch Resoluton

CEL I and T4 DNA Polymerase Concentrations Compared

The GRAMMR reaction involves the interaction of numerous enzymatic activities. Several parameters associated with the GRAMMR reaction were examined, such as CEL I concentration, T4 DNA polymerase concentration, reaction temperature, substitution of T4 DNA polymerase with T7 DNA polymerase, the presence of Taq DNA polymerase, and the source of the CEL I enzyme. A matrix of three different CEL I concentrations versus two concentrations of T4 DNA polymerase was set up to examine the limits of the *in vitro* DNA mismatch resolution reaction.

Twenty-one nanograms (21ng) of the circular double-stranded heteroduplexed plasmid, prepared as described above, was used as substrate in a series of ten microliter reactions containing 1X NEB ligase buffer, 0.5 mM each dNTP, 1.0 unit Taq DNA polymerase, 0.2 units T4 DNA ligase (Gibco/BRL), either 1.0 or 0.2 units T4 DNA polymerase, and either 0.3, 0.1, or 0.03 microliters of a CEL I preparation (fraction 5, described in Example 1). Six reactions representing all six combinations of the two T4 DNA polymerase concentrations with the three CEL I concentrations were prepared, split into equivalent sets of five microliters, and incubated at either 20 degrees C or 37 degrees C. A control reaction containing no CEL I

and 0.2 unit of T4 DNA polymerase with the other reaction components was prepared and incubated at 37 degrees C. After 30 minutes, one microliter aliquots of each reaction were transformed into competent DH5-alpha *E. coli* which were then plated on LB amp plates. Colonies were picked and cultured. Plasmid DNA was extracted and examined by restriction fragment length polymorphism analysis (RFLP) followed by sequence analysis of the GFP gene sequences. RFLP analysis was based on differences in several restriction enzyme recognition sites between the wild-type and Cycle 3 GFP genes. The RFLP results showed that throughout the CEL I /T4 DNA polymerase/temperature matrix, reassortment of restriction sites, that is GRAMMR, had occurred, and that no such reassortment had occurred in the zero CEL I control clones. DNA sequence analysis confirmed that reassortment had occurred in all of the CEL I -containing samples. Sequencing also confirmed that the zero- CEL I controls were not reassorted, with the exception of a single clone of the 16 control clones, which had a single-base change from one gene sequence to the other, presumably resulting either from repair in *E. coli* or from random mutation. The sequences of several exemplary GRAMMR-reassorted GFP clones are shown; all of which came from the reaction containing 0.3 microliters of the CEL I preparation and 1.0 unit of T4 DNA polymerase incubated at 37 degrees C. The parental wild-type and Cycle 3 GFP genes are shown first for reference.

Example 7

Taq DNA Polymerase is Not Required for Genetic Reassortment by DNA Mismatch Resolution

This experiment teaches that Taq DNA Polymerase does not dramatically, if at all, contribute or interfere with the functioning of Genetic Reassortment by DNA Mismatch Resolution (GRAMMR). Taq DNA polymerase is reported to have a 5' flap-ase activity, and had been included in the teachings of the previous examples as a safeguard against the possible formation and persistence of undesirable 5' flaps in the heteroduplexed DNA undergoing GRAMMR.

GRAMMR reactions were set up, as in Example 6, with twenty-one nanograms of the circular double-stranded heteroduplexed GFP plasmid substrate in ten microliter reactions containing 1X NEB ligase buffer, 0.5 mM each dNTP, 0.2 units T4 DNA ligase, 1.0 unit T4 DNA polymerase, 1.0 microliter of a CEL I preparation (fraction 5, described in Example 1), and either 2.5 units, 0.5 units of Taq DNA polymerase, or no Taq DNA polymerase. After 30 minutes, one microliter aliquots of each reaction were transformed into competent DH5-alpha *E. coli* which were then plated on LB amp plates. Colonies were picked and cultured. Plasmid DNA was extracted and examined by RFLP analysis followed by sequence analysis of the GFP gene sequences. The RFLP results showed that reassortment of restriction sites, that is, GRAMMR, had occurred both in the presence and

the absence of Taq DNA polymerase in the GRAMMR reaction. DNA sequence analysis confirmed these results. Therefore, the data shows that Taq DNA polymerase was unnecessary for GRAMMR.

Example 8

Alternate Proofreading DNA Polymerases for Genetic Reassortment by DNA Mismatch Resolution

This experiment teaches that Genetic Reassortment by DNA Mismatch Resolution is not limited to the use of T4 DNA polymerase, and that alternate DNA polymerases can be substituted for it.

Reactions were set up, as in Example 6, with twenty-one nanograms of the circular double-stranded heteroduplexed GFP plasmid substrate in ten microliter reactions containing 1X NEB ligase buffer, 0.5 mM each dNTP, 0.2 units T4 DNA ligase (Gibco/BRL), 10 units or 2 units of T7 DNA polymerase, 1.0 microliter of a CEL I preparation (fraction 5, described in Example 1), and 2.5 units of Taq DNA polymerase. After 30 minutes, one microliter aliquots of each reaction were transformed into competent DH5-alpha *E. coli* which were then plated on LB amp plates. Colonies were picked and cultured. Plasmid DNA was extracted and examined by RFLP analysis followed by sequence analysis of the GFP gene sequences. The RFLP results showed that reassortment of restriction sites, that is GRAMMR, had occurred in both T7 DNA polymerase-containing reactions. DNA sequence analysis

confirmed these results. Therefore, the data shows that T7 DNA polymerase can substitute for T4 DNA polymerase for GRAMMR. In addition, it shows that individual components and functionalities can be broadly substituted in GRAMMR, while still obtaining similar results.

Example 9

Use of Cloned CEL I in the GRAMMR Reaction

This example teaches that CEL I from a cloned source can be used in place of native CEL I enzyme purified from celery in Genetic Reassortment By DNA Mismatch Resolution without any noticeable change in results.

The cDNA of CEL I was cloned from celery RNA. The gene was inserted into a TMV viral vector and expressed. Transcripts of the construct were used to infect *Nicotiana benthamiana* plants. Infected tissue was harvested, and the CEL I enzyme was purified. The GRAMMR results obtained using the purified enzyme were compared to those using CEL I purified from celery, and were found to be similar.

Reactions were set up using twenty-one nanograms of the circular double-stranded heteroduplexed GFP plasmid substrate in ten microliters containing 1X NEB ligase buffer, 0.5 mM each dNTP, 0.2 units T4 DNA ligase (Gibco/BRL), 1 unit of T4 DNA polymerase, and either 1.0 microliter of CEL I purified from celery (fraction 5, described in Example 1), or 0.3 microliters

of CEL I purified from a cloned source. After 30 minutes, one microliter aliquots of each reaction were transformed into competent DH5-alpha *E. coli* which were then plated on LB amp plates. Colonies were picked and cultured. Plasmid DNA was extracted and examined by RFLP analysis followed by sequence analysis of the GFP gene sequences. The RFLP results showed that reassortment of restriction sites, that is, GRAMMR had occurred in both celery-derived CEL I, as well as cloned CEL I-containing reactions. DNA sequence analysis confirmed these results. Therefore, the data shows CEL I from a cloned source can be used in lieu of CEL I from celery for GRAMMR. In addition, the data demonstrates that it is CEL I activity that is part of the GRAMMR method, rather than a coincidental effect resulting from the purifying steps used in extracting CEL I from celery.

Example 10

Molecular Breeding of Tobamovirus 30K Genes in a Viral Vector.

In the preceding examples, Genetic Reassortment by DNA Mismatch Resolution has been taught to be useful for reassorting sequences that are highly homologous, for example, wtGFP and Cycle 3 GFP are 96% identical. The present example teaches that GRAMMR can be used to reassort more divergent nucleic acid sequences, such as genes encoding tobamovirus movement protein genes.

Heteroduplexes of two tobamovirus movement protein (MP) genes that are approximately 75% identical

were generated. The heteroduplex substrate was prepared by annealing partially-complementary single-stranded DNAs of opposite strandedness synthesized by asymmetric PCR; one strand encoding the movement protein gene from the tobacco mosaic virus U1 type strain (TMV-U1) (SEQ ID NO:09), and the other strand encoding the movement protein gene from tomato mosaic virus (ToMV) (SEQ ID NO:10). The sequences of the two partially complementary movement protein genes were flanked by 33 nucleotides of absolute complementarity to promote annealing of the DNAs at their termini and to facilitate PCR amplification and cloning. The annealing reaction took place by mixing 2.5 micrograms of each single-stranded DNA in a 150 microliter reaction containing 333 mM NaCl, 33 mM MgCl2, 3.3 mM dithiothreitol, 166mM Tris-HCl, pH 7, and incubating at 95°C for one minute followed by slow cooling to room-temperature. GRAMMR was performed by incubating 5 microliters of the heteroduplex substrate in a 20 microliter reaction containing 1X NEB ligase buffer, 0.5 mM each dNTP, 0.4 units T4 DNA ligase (Gibco/BRL), 2.0 units of T4 DNA polymerase, and CEL I. The CEL I was from a cloned preparation and the amount that was used varied from 2 microliters of the prep, followed by five serial 3-fold dilutions. A seventh preparation with no CEL I was prepared, which served as a control.

After one hour at room-temperature, DNA was purified from the reactions using Strataprep spin DNA purification columns (Stratagene, LaJolla, CA) and used as templates for PCR reactions using primers designed to anneal to the flanking primer-binding sites of the two

sequences. PCR products from each reaction were purified using Strataprep columns, digested with *Avr*II and *Pac*I, and ligated into the movement protein slot of similarly-cut pGENEWARE-MP-Avr-Pac. This plasmid contained a full-length infectious tobamovirus-GFP clone modified with *Avr*II and *Pac*I sites flanking the movement protein gene to permit its replacement by other movement protein genes. After transformation of DH5-alpha *E. coli* and plating, colonies were picked, cultures grown, and DNA was extracted. The movement protein inserts were subjected to DNA sequence analysis from both directions and the sequence data confirmed that in the majority of inserts derived from the GRAMMR-treated material were reassorted sequences made up of both TMV-U1 and ToMV movement protein gene sequences. The DNA sequences of several exemplary GRAMMR MP clones are shown as SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:14, and SEQ ID NO:15.

EXAMPLE 11

GRAMMR Reassortment to Generate Improved Arsenate Detoxifying Bacteria

Arsenic detoxification is important for mining of arsenopyrite-containing gold ores and other uses, such as environmental remediation. Plasmid pGJ103, containing an arsenate detoxification operon (Ji and Silver, 1992)(Ji, G. and Silver, S., Regulation and expression of the arsenic resistance operon from Staphylococcus aureus plasmid pI258, J. Bacteriol. 174, 3684-3694 (1992), incorporated herein by reference), is obtained from Prof.

Simon Silver (U. of Illinois, Chicago, Ill.). *E. coli* TG1 containing pGJ103, containing the pI258 *ars* operon cloned into pUC19, has a MIC (minimum inhibitory concentration) of 4 µg/ml on LB ampicillin agar plates. The *ars* operon is amplified by mutagenic PCR [REF], cloned into pUC19, and transformed into *E. coli* TG1. Transformed cells are plated on a range of sodium arsenate concentrations (2, 4, 8, 16 mM). Colonies from the plates with the highest arsenate levels are picked. The colonies are grown in a mixed culture with appropriate arsenate selection. Plasmid DNA is isolated from the culture. The plasmid DNA is linearized by digestion with a restriction endonuclease that cuts once into the pUC19 plasmid backbone. The linearized plasmids are denatured by heating 10 min. at 94°C. The reaction is allowed to cool to promote annealing of the single strands. Partially complementary strands that hybridize have non-basepaired nucleotides at the sites of the mismatches. Treatment with CEL I (purified by the method of Example 9) causes nicking of one or the other polynucleotide strand 3' of each mismatch. The presence of a polymerase containing a 3'-to-5' exonuclease ("proofreading") activity, such as T4 DNA polymerase allows excision of the mismatch, and subsequent 5'-to-3' polymerase activity fills in the gap using the other strand as a template. T4 DNA ligase then seals the nick by restoring the phosphate backbone of the repaired strand. The result is a randomization of mutations among input strands to give output strands with potentially improved properties. These output polynucleotides are transformed directly into *E. coli* TG1 and the cells are

plated at higher arsenate levels; 8, 16, 32, 64 mM. Colonies are picked from the plates with the highest arsenate levels and another round of reassortment is performed as above except that resulting transformed cells are plated at 32, 64, 128, 256 mM arsenate. The process can then be repeated one or more times with the selected clones in an attempt to obtain additional improvements.

What is claimed is:

1. An *in vitro* method of making sequence variants from at least one heteroduplex polynucleotide where said heteroduplex has at least two non-complementary nucleotide base pairs, said method comprising:

a. preparing at least one heteroduplex polynucleotide;

b. combining said heteroduplex polynucleotide with an effective amount of CEL I, T4 DNA polymerase, and T4 DNA ligase; and

c. allowing sufficient time for the percentage of complementarity to increase, wherein one or more variants are made.

2. An *in vitro* method of making sequence variants from at least one heteroduplex polynucleotide wherein said heteroduplex has at least two non-complementary nucleotide base pairs, said method comprising:

a. preparing at least one heteroduplex polynucleotide;

b. combining said heteroduplex polynucleotide with an effective amount of an agent or agents with exonuclease activity, polymerase activity and strand cleavage activity; and

c. allowing sufficient time for the percentage of complementarity to increase,

wherein at least one or more variants are made.

3. The method of claim 2 wherein said heteroduplex polynucleotide is circular.

4. The method of claim 2 wherein said heteroduplex polynucleotide is linear.

5. The method of claim 3 or 4 wherein said heteroduplex polynucleotide is a replicon.

6. The method of claim 2 wherein said variants have different amounts of complementarity.

7. The method of claim 2 wherein said agents having exonuclease activity, polymerase activity, and strand cleavage activity are added sequentially.

8. The method of claim 2 wherein said agents having exonuclease activity, polymerase activity, and strand cleavage activity are added concurrently.

9. The method of claim 2 in step (b) further comprising ligase activity.

10. The method of claim 2 further comprising a step of,

(d) adding a ligase.

11. The method of claim 2 wherein said agents having exonuclease activity, polymerase activity, ligase activity, and strand cleavage activity are added sequentially.

12. The method of claim 2 wherein said agents having exonuclease activity, polymerase activity, ligase activity, and strand cleavage activity are added concurrently.

13. The method of claim 9 wherein said ligase is T4 DNA ligase, E. coli DNA ligase, or Taq DNA ligase.

14. The method of claim 2 wherein said agent with strand cleavage activity is an enzyme.

15. The method of claim 2 wherein said agent with strand cleavage activity is a mismatch endonuclease.

16. The method of claim 2 wherein said agent with strand cleavage activity is selected from the group consisting of CEL I, T4 endonuclease VII, T7 endonuclease I, S1 nuclease, BAL-31 nuclease, FEN1, cleavase, pancreatic DNase I, SP nuclease, mung bean nuclease, and nuclease P1.

17. The method of claim 2 wherein said agent with strand cleavage activity is a chemical.

18. The method of claim 2 wherein said agent with strand cleavage activity is selected from the group consisting of potassium permanganate, tetraethylammonium acetate, sterically bulky photoactivatable DNA intercalators, [Rh(bpy)2(chrysi)]3+, osmium tetroxide with piperidine, and hydroxylamine with piperidine.

19. The method of claim 2 wherein said agent with strand cleavage activity is ionizing radiation, or kinetic radiation.

20. The method of claim 2 wherein said agent with polymerase activity is T4 DNA polymerase.

21. The method of claim 2 wherein said agent with polymerase activity is T7 DNA polymerase.

22. The method of claim 2 wherein said agent with both polymerase activity and 3' to 5' exonuclease activity is T4 DNA polymerase, T7 DNA polymerase, E. coli Pol 1, or Pfu DNA polymerase.

23. The method of claim 2 wherein said agent with both polymerase activity and 5' to 3' exonuclease activity is E. coli Pol 1.

24. The method of claim 2 wherein said effective amount of strand cleavage activity, and exonuclease activity/polymerase activity and ligase activity are provided by CEL I, T4 DNA polymerase, and T4 DNA ligase.

25. The method of claim 2 wherein said effective amount of strand cleavage activity, and exonuclease activity/polymerase activity and ligase activity are provided by CEL I, T7 DNA polymerase, and T4 DNA ligase.

26. The method of claim 2 wherein an effective amount of strand cleavage activity, and exonuclease activity/polymerase activity and ligase activity are provided by T4 endonuclease VII, T4 DNA polymerase, and T4 DNA ligase.

27. An in vitro method of increasing diversity in a population of sequences, comprising: preparing at least one heteroduplex polynucleotide; combining said heteroduplex polynucleotide with an effective amount of an agent or agents with 3' to 5' exonuclease activity, polymerase activity and strand cleavage activity; and allowing sufficient time for the percentage of complementarity to increase, wherein diversity in the population is increased.

28. The method of claim 27 wherein said heteroduplex polynucleotide is circular.

29. The method of claim 27 wherein said heteroduplex polynucleotide is linear.

30. The method of claim 3 or 4 wherein said heteroduplex polynucleotide is a replicon.

31. The method of claim 27 wherein said variants have different amounts of complementarity.

32. The method of claim 27 wherein said enzymes having 3' to 5' exonuclease activity, polymerase activity, and strand cleavage activity are added sequentially.

33. The method of claim 27 wherein said enzymes having 3' to 5' exonuclease activity, polymerase activity, and strand cleavage activity are added at the same time.

34. The method of claim 27 further comprising adding a ligase.

35. The method of claim 9 wherein said ligase is T4 DNA ligase, E. coli DNA ligase, or Taq DNA ligase.

36. The method of claim 27 wherein said agent with strand cleavage activity is an enzyme.

37. The method of claim 27 wherein said agent with strand cleavage activity is a mismatch endonuclease.

38. The method of claim 27 wherein said agent with strand cleavage activity is selected from the group consisting of CEL I, T4 endonuclease VII, T7 endonuclease I, S1 nuclease, BAL-31 nuclease, FEN1, cleavase, pancreatic DNase I, SP nuclease, mung bean nuclease, nuclease P1.

39. The method of claim 27 wherein said agent with strand cleavage activity is a chemical.

40. The method of claim 27 wherein said agent with strand cleavage activity is selected from the group consisting of potassium permanganate, tetraethylammonium acetate, sterically bulky photoactivatable DNA intercalators, [Rh(bpy)2(chrysi)]3+, osmium tetroxide with piperidine, and hydroxylamine with piperidine.

41. The method of claim 27 wherein said agent with strand cleavage activity is ionizing radiation, or kinetic radiation.

42. The method of claim 27 wherein said agent with polymerase activity is T4 DNA polymerase.

43. The method of claim 27 wherein said agent with polymerase activity is T7 DNA polymerase.

44. The method of claim 27 wherein said agent with both polymerase activity and 3' to 5' exonuclease activity is T4 DNA polymerase, T7 DNA polymerase, E. coli Pol 1, or Pfu DNA polymerase.

45. The method of claim 27 wherein said agent with both polymerase activity and 5' to 3' exonuclease activity is E. coli Pol 1.

46. The method of claim 27 wherein said effective amount of strand cleavage activity, exonuclease activity/polymerase activity and ligase activity are provided by CEL I, T4 DNA polymerase, and T4 DNA ligase respectively.

47. The method of claim 27 wherein said effective amount of strand cleavage activity, and exonuclease activity/polymerase activity and ligase activity are provided by CEL I, T7 DNA polymerase, and T4 DNA ligase respectively.

48. The method of claim 27 wherein said effective amount of strand cleavage activity, and exonuclease activity/polymerase activity and ligase activity are provided by CEL I, T7 DNA polymerase, and T4 DNA ligase respectively.

49. The method of claim 27 wherein an effective amount of strand cleavage activity, and exonuclease activity/polymerase activity and ligase activity are provided by T4 endonuclease VII, T4 DNA polymerase, and T4 DNA ligase respectively.

50. An in vitro method of increasing diversity in a population of sequences, comprising:

a. preparing at least one heteroduplex polynucleotide;

b. combining said heteroduplex polynucleotide with an effective amount of CEL I, T4 DNA polymerase, and T4 DNA ligase; and

c. allowing sufficient time for the percentage of complementarity to increase, wherein diversity in the population is increased.

51. A method of obtaining a polynucleotide encoding a desired functional property, comprising:

a. preparing at least one heteroduplex polynucleotide;

b. combining said heteroduplex polynucleotide with an effective amount of an agent or agents with exonuclease activity, polymerase activity, and strand cleavage activity;

c. allowing sufficient time for the percentage of complementarity between strands of the heteroduplex polynucleotide

to increase, wherein diversity in the population is increased; and

d. screening or selecting a population of variants for the desired functional property.

52. A method of obtaining a polynucleotide encoding a desired functional property, comprising:

a. preparing at least one heteroduplex polynucleotide;

b. combining said heteroduplex polynucleotide with an effective amount of an agent or agents with exonuclease activity, polymerase activity, and strand cleavage activity;

c. allowing sufficient time for the percentage of complementarity between strands of the heteroduplex polynucleotide to increase, wherein diversity in the population is increased;

d. converting DNA to RNA; and

e. screening or selecting a population of ribonucleic acid variants for the desired functional property.

53. A method of obtaining a polypeptide having a desired functional property, comprising:

a. preparing at least one heteroduplex polynucleotide;

b. combining said heteroduplex polynucleotide with an effective amount of an agent or

agents with exonuclease activity, polymerase activity and strand cleavage activity;

c. allowing sufficient time for the percentage of complementarity between strands of said heteroduplex polynucleotide to increase,

d. converting said heteroduplex polynucleotide to RNA, and said RNA to a polypeptide; and

e. and screening or selecting a population of polypeptide variants for said desired functional property.

54. A method of obtaining a polynucleotide encoding a desired functional property, comprising:

a. preparing at least one heteroduplex polynucleotide;

b. combining said heteroduplex polynucleotide with an effective amount of an agent or agents with exonuclease activity, polymerase activity and strand cleavage activity;

c. allowing sufficient time for the percentage of complementarity between strands of said heteroduplex polynucleotide to increase,

d. screening or selecting for a population of variants having a desired functional property;

e. denaturing said population of variants to obtain single strand polynucleotides;

f. annealing said single strand polynucleotides to form at least one second heteroduplex polynucleotide;

g. combining said second heteroduplex polynucleotide with an effective amount of an agent or agents with exonuclease activity, polymerase activity and strand cleavage activity; and

h. allowing sufficient time for the percentage of complementarity between strands of the heteroduplex polynucleotide to increase.

55. The method of claim 54 wherein said heteroduplex polynucleotide is greater than 95% identical.

56. The method of claim 54 wherein said heteroduplex polynucleotide is about 95% identical.

57. The method of claim 54 wherein said heteroduplex polynucleotide is about 90% identical.

58. The method of claim 54 wherein said heteroduplex polynucleotide is about 85% identical.

59. The method of claim 54 wherein said heteroduplex polynucleotide is about 80% identical.

60. The method of claim 54 wherein said heteroduplex polynucleotide is about 75% identical.

61. The method of claim 2 wherein the heteroduplex polynucleotide is about 1000Kb.

62. The method of claim 2 wherein the heteroduplex polynucleotide is about 10,000Kb.

63. The method of claim 2 wherein the heteroduplex polynucleotide is about 100,000Kb.

64. A kit used for increasing diversity in a population of sequences, comprising: preparing at least one heteroduplex polynucleotide; combining said heteroduplex polynucleotide with an effective amount of an agent or agents with 3' to 5' exonuclease activity, polymerase activity and strand cleavage activity; and allowing sufficient time for the percentage of complementarity to increase, wherein diversity in the population is increased.

65. The kit of claim 64 further comprising having a ligase activity.

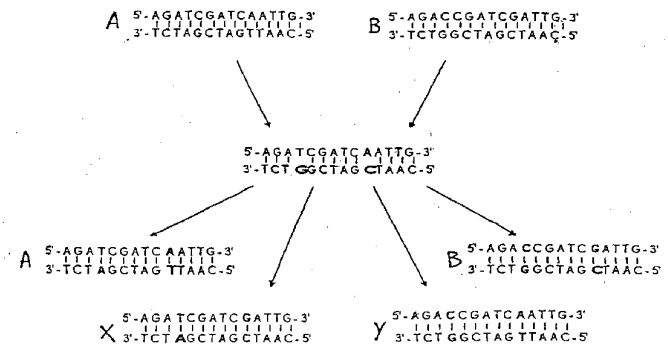

A 5'-AGATCGATCAATTG-3'
3'-TCTAGCTAGTTAAC-5'
B 5'-AGACCGATCGATTG-3'
3'-TCTGGCTAGCTAAC-5'
5'-AGATCGATCAATTG-3'
3'-TCTGGCTAGCTAAC-5'
A 5'-AGATCGATCAATTG-3'
3'-TCTAGCTAGTTAAC-5'
B 5'-AGACCGATCGATTG-3'
3'-TCTGGCTAGCTAAC-5'
X 5'-AGATCGATCGATTG-3'
3'-TCTAGCTAGCTAAC-5'
Y 5'-AGACCGATCAATTG-3'
3'-TCTGGCTAGTTAAC-5'

FIGURE 1

FIG. 2A



FIG. 2B

POSSIBLE STRAND COMBINATIONS	POSSIBLE +/- STRAND COMBINATIONS	PARTIALLY COMPLEMENTARY POPULATIONS
1+/2-	✓	
1+/3+		
1+/4-	✓	✓
2-/3+	✓	✓
2-/4-		
3+/4-	✓	

FIGURE 2

SEQUENCE LISTING

<110> Large Scale Biology Corporation

<120> A Method of Increasing Complementarity
In A Heteroduplex

<130> FP-LG 5058

<150> US 60/266,386
<151> 2001-02-02

<150> US 60/268,785
<151> 2001-02-14

<160> 15

<170> FastSEQ for Windows Version 4.0

<210> 1
<211> 717
<212> DNA
<213> Aequorea victoria

<400> 1
atgagtaaag gagaagaact tttcactgga gttgtcccaa ttcttgttga attagatggt 60
gatgttaatg ggcacaaatt ttctgtcagt ggagagggtg aaggtgatgc aacatacgga 120
aaacttaccc ttaaatttat ttgcactact ggaaaactac ctgttccatg gccaacactt 180
gtcactactt tctcttatgg tgttcaatgc ttttcaagat acccagatca tatgaaacgg 240
catgactttt tcaagagtgc catgcccgaa ggttatgtac aggaaagaac tatattttc 300
aaggatgacg ggaactacaa gacacgtgct gaagtcaagt ttgaaggtga tacccttgtt 360
aatagaatcg agttaaaagg tattgatttt aaagaagatg gaaacattct tggacacaaa 420
ttggaataca actataactc acacaatgta tacatcatgg cagacaaaca aaagaatgga 480
atcaaagtta acttcaaaat tagacacaac attgaagatg gaagcgttca actagcagac 540
cattatcaac aaaatactcc aattggcgat ggccctgtcc ttttaccaga caaccattac 600
ctgtccacac aatctgccct ttcgaaagat cccaacgaaa agagagacca catggtcctt 660
cttgagtttg taacagctgc tgggattaca catggcatgg atgaactata caaataa 717

<210> 2
<211> 717
<212> DNA
<213> Artificial Sequence

<220>
<223> synthetic construct

<400> 2
atgagtaaag gagaagaact ttcactgga gttgtcccaa ttcttgttga attagatggt 60
gatgttaatg ggcacaaatt ttctgtcagt ggagagggtg aaggtgatgc tacatacgga 120
aagcttaccc ttaaatttat ttgcactact ggaaaactac ctgttccatg gccaacactt 180
gtcactactt tctcttatgg tgttcaatgc ttttcccgtt atccggatca tatgaaacgg 240
catgactttt tcaagagtgc catgcccgaa ggttatgtac aggaacgcac tatatctttc 300
aaagatgacg ggaactacaa gacgcgtgct gaagtcaagt ttgaaggtga tacccttgtt 360
aatcgtatcg agttaaaagg tattgatttt aaagaagatg gaaacattct cggacacaaa 420

ctcgagtaca actataactc acacaatgta tacatcacgg cagacaaaca aaagaatgga 480
atcaaagcta acttcaaaat tcgccacaac attgaagatg gatccgttca actagcagac 540
cattatcaac aaaatactcc aattggcgat ggccctgtcc ttttaccaga caaccattac 600
ctgtcgacac aatctgccct ttcgaaagat cccaacgaaa agcgtgacca catggtcctt 660
cttgagtttg taactgctgc tgggattaca catggcatgg atgaactata caaataa 717

<210> 3
<211> 3637
<212> DNA
<213> Artificial Sequence

<220>
<223> synthetic construct

<400> 3
gtggcacttt tcggggaaat gtgcgcggaa cccctatttg tttatttttc taaatacatt 60
caaatatgta tccgctcatg agacaataac cctgataaat gcttcaataa tattgaaaaa 120
ggaagagtat gagtattcaa catttccgtg tcgcccttat tcccttttt gcggcatttt 180
gccttcctgt ttttgctcac ccagaaacgc tggtgaaagt aaaagatgct gaagatcagt 240
tgggtgcacg agtgggttac atcgaactgg atctcaacag cggtaagatc cttgagagtt 300
ttcgccccga agaacgtttt ccaatgatga gcacttttaa agttctgcta tgtggcgcgg 360
tattatcccg tattgacgcc gggcaagagc aactcggtcg ccgcatacac tattctcaga 420
atgacttggt tgagtactca ccagtcacag aaaagcatct tacggatggc atgacagtaa 480
gagaattatg cagtgctgcc ataaccatga gtgataacac tgcggccaac ttacttctga 540
caacgatcgg aggaccgaag gagctaaccg ctttttgca caacatgggg gatcatgtaa 600
ctcgccttga tcgttgggaa ccggagctga atgaagccat accaaacgac gagcgtgaca 660
ccacgatgcc tgtagcaatg gcaacaacgt tgcgcaaact attaactggc gaactactta 720
ctctagcttc ccggcaacaa ttaatagact ggatggaggc ggataaagtt gcaggaccac 780
ttctgcgctc ggcccttccg gctggctggt ttattgctga taaatctgga gccggtgagc 840
gtgggtctcg cggtatcatt gcagcactgg ggccagatgg taagccctcc cgtatcgtag 900
ttatctacac gacggggagt caggcaacta tggatgaacg aaatagacag atcgctgaga 960
taggtgcctc actgattaag cattggtaac tgtcagacca agtttactca tatatacttt 1020
agattgattt aaaacttcat ttttaattta aaaggatcta ggtgaagatc ctttttgata 1080
atctcatgac caaaatccct taacgtgagt tttcgttcca ctgagcgtca gaccccgtag 1140
aaaagatcaa aggatcttct tgagatcctt ttttctgcg cgtaatctgc tgcttgcaaa 1200
caaaaaaacc accgctacca gcggtggttt gtttgccgga tcaagagcta ccaactcttt 1260
ttccgaaggt aactggcttc agcagagcgc agataccaaa tactgtcctt ctagtgtagc 1320
cgtagttagg ccaccacttc aagaactctg tagcaccgcc tacatacctc gctctgctaa 1380
tcctgttacc agtggctgct gccagtggcg ataagtcgtg tcttaccggg ttggactcaa 1440
gacgatagtt accggataag gcgcagcggt cgggctgaac gggggttcg tgcacacagc 1500
ccagcttgga gcgaacgacc tacaccgaac tgagatacct acagcgtgag ctatgagaaa 1560
gcgccacgct tcccgaaggg agaaaggcgg acaggtatcc ggtaagcggc agggtcggaa 1620
caggagagcg cacgagggag cttccagggg gaaacgcctg gtatctttat agtcctgtcg 1680
ggtttcgcca cctctgactt gagcgtcgat ttttgtgatg ctcgtcaggg gggcggagcc 1740
tatggaaaaa cgccagcaac gcggcctttt tacggttcct ggccttttgc tggccttttg 1800
ctcacatgtt ctttcctgcg ttatcccctg attctgtgga taaccgtatt accgcctttg 1860
agtgagctga taccgctcgc cgcagccgaa cgaccgagcg cagcgagtca gtgagcgagg 1920
aagcggaaga gcgcccaata cgcaaaccgc ctctccccgc gcgttggccg attcattaat 1980
gcagctggca cgacaggttt cccgactgga aagcgggcag tgagcgcaac gcaattaatg 2040
tgagttagct cactcattag gcaccccagg ctttacactt tatgcttccg gctcgtatgt 2100
tgtgtggaat tgtgagcgga taacaatttc acacaggaaa cagctatgac catgattacg 2160
ccaagcgcgc aattaaccct cactaaaggg aacaaaagct gggtaccgat gagtaaagga 2220
gaagaacttt tcactggagt tgtcccaatt cttgttgaat tagatggtga tgttaatggg 2280
cacaaatttt ctgtcagtgg agagggtgaa ggtgatgcaa catacggaaa acttaccctt 2340
aaatttattt gcactactgg aaaactacct gttccatggc caacacttgt cactactttc 2400

tcttatggtg ttcaatgctt ttcaagatac ccagatcata tgaaacggca tgacttttte 2460
aagagtgcca tgcccgaagg ttatgtacag gaaagaacta tatttttcaa ggatgacggg 2520
aactacaaga cacgtgctga agtcaagttt gaaggtgata cccttgttaa tagaatcgag 2580
ttaaaaggta ttgattttaa agaagatgga aacattcttg gacacaaatt ggaatacaac 2640
tataactcac acaatgtata catcatggca gacaaacaaa agaatggaat caaagttaac 2700
ttcaaaatta gacacaacat tgaagatgga agcgttcaac tagcagacca ttatcaacaa 2760
aatactccaa ttggcgatgg ccctgtcctt ttaccagaca accattacct gtccacacaa 2820
tctgcccttt cgaaagatcc caacgaaaag agagaccaca tggtccttct tgagtttgta 2880
acagctgctg ggattacaca tggcatggat gaactataca aataagaatt cctgcagccc 2940
gggggatcca ctagttctag agcggccgcc accgcggtgg agctccaatt cgccctatag 3000
tgagtcgtat tacgcgcgct cactggccgt cgttttacaa cgtcgtgact gggaaaaccc 3060
tggcgttacc caacttaatc gccttgcagc acatcccccc ttcgccagct ggcgtaatag 3120
cgaagaggcc cgcaccgatc gcccttccca acagttgcgc agcctgaatg gcgaatggga 3180
cgcgccctgt agcggcgcat taagcgcggc gggtgtggtg gttacgcgca gcgtgaccgc 3240
tacacttgcc agcgccctag cgcccgctcc tttcgctttc ttcccttcct ttctcgccac 3300
gttcgccggc tttccccgtc aagctctaaa tcgggggctc cctttagggt tccgatttag 3360
tgctttacgg cacctcgacc ccaaaaaact tgattagggt gatggttcac gtagtgggcc 3420
atcgccctga tagacggttt ttcgcccttt gacgttggag tccacgttct ttaatagtgg 3480
actcttgttc caaactggaa caacactcaa ccctatctcg gtctattctt ttgatttata 3540
agggattttg ccgatttcgg cctattggtt aaaaaatgag ctgatttaac aaaaatttaa 3600
cgcgaatttt aacaaaatat taacgcttac aatttag 3637

<210> 4
<211> 3637
<212> DNA
<213> Artificial Sequence

<220>
<223> synthetic construct

<400> 4
gtggcacttt tcggggaaat gtgcgcggaa cccctatttg tttatttttc taaatacatt 60
caaatatgta tccgctcatg agacaataac cctgataaat gcttcaataa tattgaaaaa 120
ggaagagtat gagtattcaa catttccgtg tcgcccttat tccctttttt gcggcatttt 180
gccttcctgt ttttgctcac ccagaaacgc tggtgaaagt aaaagatgct gaagatcagt 240
tgggtgcacg agtgggttac atcgaactgg atctcaacag cggtaagatc cttgagagtt 300
ttcgccccga agaacgtttt ccaatgatga gcacttttaa agttctgcta tgtggcgcgg 360
tattatcccg tattgacgcc gggcaagagc aactcggtcg ccgcatacac tattctcaga 420
atgacttggt tgagtactca ccagtcacag aaaagcatct tacggatggc atgacagtaa 480
gagaattatg cagtgctgcc ataaccatga gtgataacac tgcggccaac ttacttctga 540
caacgatcgg aggaccgaag gagctaaccg ctttttttgca caacatgggg gatcatgtaa 600
ctcgccttga tcgttgggaa ccggagctga atgaagccat accaaacgac gagcgtgaca 660
ccacgatgcc tgtagcaatg gcaacaacgt tgcgcaaact attaactggc gaactactta 720
ctctagcttc ccggcaacaa ttaatagact ggatggaggc ggataaagtt gcaggaccac 780
ttctgcgctc ggcccttccg gctggctggt ttattgctga taaatctgga gccggtgagc 840
gtgggtctcg cggtatcatt gcagcactgg ggccagatgg taagccctcc cgtatcgtag 900
ttatctacac gacggggagt caggcaacta tggatgaacg aaatagacag atcgctgaga 960
taggtgcctc actgattaag cattggtaac tgtcagacca agtttactca tatatacttt 1020
agattgattt aaaacttcat ttttaattta aaaggatcta ggtgaagatc ctttttgata 1080
atctcatgac caaaatccct taacgtgagt tttcgttcca ctgagcgtca gaccccgtag 1140
aaaagatcaa aggatcttct tgagatcctt ttttctgcg cgtaatctgc tgcttgcaaa 1200
caaaaaaacc accgctacca gcggtggttt gtttgccgga tcaagagcta ccaactcttt 1260
ttccgaaggt aactggcttc agcagagcgc agataccaaa tactgtcctt ctagtgtagc 1320
cgtagttagg ccaccacttc aagaactctg tagcaccgcc tacatacctc gctctgctaa 1380
tcctgttacc agtggctgct gccagtggcg ataagtcgtg tcttaccggg ttggactcaa 1440

gacgatagtt accggataag gcgcagcggt cgggctgaac gggggttcg tgcacacagc 1500
ccagcttgga gcgaacgacc tacaccgaac tgagatacct acagcgtgag ctatgagaaa 1560
gcgccacgct tcccgaaggg agaaaggcgg acaggtatcc ggtaagcggc agggtcggaa 1620
caggagagcg cacgagggag cttccagggg gaaacgcctg gtatctttat agtcctgtcg 1680
ggtttcgcca cctctgactt gagcgtcgat ttttgtgatg ctcgtcaggg gggcggagcc 1740
tatggaaaaa cgccagcaac gcggcctttt tacggttcct ggccttttgc tggccttttg 1800
ctcacatgtt ctttcctgcg ttatcccctg attctgtgga taaccgtatt accgcctttg 1860
agtgagctga taccgctcgc cgcagccgaa cgaccgagcg cagcgagtca gtgagcgagg 1920
aagcggaaga gcgcccaata cgcaaaccgc ctctccccgc gcgttggccg attcattaat 1980
gcagctggca cgacaggttt cccgactgga aagcgggcag tgagcgcaac gcaattaatg 2040
tgagttagct cactcattag gcaccccagg ctttacactt tatgcttccg gctcgtatgt 2100
tgtgtggaat tgtgagcgga taacaatttc acacaggaaa cagctatgac catgattacg 2160
ccaagcgcgc aattaaccct cactaaaggg aacaaaagct gggtaccgat gagtaaagga 2220
gaagaacttt tcactggagt tgtcccaatt cttgttgaat tagatggtga tgttaatggg 2280
cacaaatttt ctgtcagtgg agagggtgaa ggtgatgcta catacggaaa gcttaccctt 2340
aaatttattt gcactactgg aaaactacct gttccatggc caacacttgt cactactttc 2400
tcttatggtg ttcaatgctt ttcccgttat ccggatcata tgaaacggca tgacttttc 2460
aagagtgcca tgccgaagg ttatgtacag gaacgcacta tatctttcaa agatgacggg 2520
aactacaaga cgcgtgctga agtcaagttt gaaggtgata cccttgttaa tcgtatcgag 2580
ttaaaaggta ttgattttaa agaagatgga aacattctcg gacacaaact cgagtacaac 2640
tataactcac acaatgtata catcacggca gacaaacaaa agaatggaat caaagctaac 2700
ttcaaaattc gccacaacat tgaagatgga tccgttcaac tagcagacca ttatcaacaa 2760
aatactccaa ttggcgatgg ccctgtcctt ttaccagaca accattacct gtcgacacaa 2820
tctgcccttt cgaaagatcc caacgaaaag cgtgaccaca tggtccttct tgagtttgta 2880
actgctgctg ggattacaca tggcatggat gaactataca aataagaatt cctgcagccc 2940
gggggatcca ctagttctag agcggccgcc accgcggtgg agctccaatt cgccctatag 3000
tgagtcgtat tacgcgcgct cactggccgt cgttttacaa cgtcgtgact gggaaaaccc 3060
tggcgttacc caacttaatc gccttgcagc acatcccccc ttcgccagct ggcgtaatag 3120
cgaagaggcc cgcaccgatc gcccttccca acagttgcgc agcctgaatg gcgaatggga 3180
cgcgccctgt agcggcgcat taagcgcggc gggtgtggtg gttacgcgca gcgtgaccgc 3240
tacacttgcc agcgccctag cgcccgctcc tttcgctttc ttcccttcct ttctcgccac 3300
gttcgccggc tttccccgtc aagctctaaa tcgggggctc cctttagggt tccgatttag 3360
tgctttacgg cacctcgacc ccaaaaaact tgattagggt gatggttcac gtagtgggcc 3420
atcgccctga tagacggttt ttcgcccttt gacgttggag tccacgttct ttaatagtgg 3480
actcttgttc caaactggaa caacactcaa ccctatctcg gtctattctt ttgatttata 3540
agggattttg ccgatttcgg cctattggtt aaaaaatgag ctgatttaac aaaaatttaa 3600
cgcgaatttt aacaaaatat taacgcttac aatttag 3637

<210> 5
<211> 717
<212> DNA
<213> Artificial Sequence

<220>
<223> synthetic construct

<400> 5
atgagtaaag gagaagaact tttcactgga gttgtcccaa ttcttgttga attagatggt 60
gatgttaatg ggcacaaatt ttctgtcagt ggagagggtg aaggtgatgc aacatacgga 120
aaacttaccc ttaaatttat ttgcactact ggaaaactac ctgttccatg gccaacactt 180
gtcactactt tctcttatgg tgttcaatgc ttttcaagat acccagatca tatgaaacgg 240
catgactttt tcaagagtgc catgcccgaa ggttatgtac aggaacgcac tatattttc 300
aaggatgacg ggaactacaa gacacgtgct gaagtcaagt ttgaaggtga tacccttgtt 360
aatagaatcg agttaaaagg tattgatttt aaagaagatg gaaacattct tggacacaaa 420
ttggaataca actataactc acacaatgta tacatcatgg cagacaaaca aaagaatgga 480

atcaaagtta acttcaaaat tagacacaac attgaagatg gaagcgttca actagcagac 540
cattatcaac aaaatactcc aattggcgat ggccctgtcc ttttaccaga caaccattac 600
ctgtccacac aatctgccct ttcgaaagat cccaacgaaa agagagacca catggtcctt 660
cttgagtttg taacagctgc tgggattaca catggcatgg atgaactata caaataa 717

<210> 6
<211> 717
<212> DNA
<213> Artificial Sequence

<220>
<223> synthetic construct

<400> 6
atgagtaaag gagaagaact tttcactgga gttgtcccaa ttcttgttga attagatggt 60
gatgttaatg ggcacaaatt ttctgtcagt ggagagggtg aaggtgatgc tacatacgga 120
aagcttaccc ttaaatttat ttgcactact ggaaaactac ctgttccatg gccaacactt 180
gtcactactt tctcttatgg tgttcaatgc ttttcaagat acccagatca tatgaaacgg 240
catgactttt tcaagagtgc catgcccgaa ggttatgtac aggaacgcac tatatctttc 300
aaagatgacg ggaactacaa gacacgtgct gaagtcaagt ttgaaggtga tacccttgtt 360
aatagaatcg agttaaaagg tattgatttt aaagaagatg gaaacattct tggacacaaa 420
ctcgagtaca actataactc acacaatgta tacatcatgg cagacaaaca aaagaatgga 480
atcaaagtta acttcaaaat tagacacaac attgaagatg gaagcgttca actagcagac 540
cattatcaac aaaatactcc aattggcgat ggccctgtcc ttttaccaga caaccattac 600
ctgtccacac aatctgccct ttcgaaagat cccaacgaaa agagagacca catggtcctt 660
cttgagtttg taacagctgc tgggattaca catggcatgg atgaactata caaataa 717

<210> 7
<211> 717
<212> DNA
<213> Artificial Sequence

<220>
<223> synthetic construct

<400> 7
atgagtaaag gagaagaact tttcactgga gttgtcccaa ttcttgttga attagatggt 60
gatgttaatg ggcacaaatt ttctgtcagt ggagagggtg aaggtgatgc tacatacgga 120
aagcttaccc ttaaatttat ttgcactact ggaaaactac ctgttccatg gccaacactt 180
gtcactactt tctcttatgg tgttcaatgc ttttcccgtt atccggatca tatgaaacgg 240
catgactttt tcaagagtgc catgcccgaa ggttatgtac aggaacgcac tatatctttc 300
aaagatgacg ggaactacaa gacgcgtgct gaagtcaagt ttgaaggtga tacccttgtt 360
aatagaatcg agttaaaagg tattgatttt aaagaagatg gaaacattct cggacacaaa 420
ttggaataca actataactc acacaatgta tacatcacgg cagacaaaca aaagaatgga 480
atcaaagcta acttcaaaat tcgccacaac attgaagatg gatccgttca actagcagac 540
cattatcaac aaaatactcc aattggcgat ggccctgtcc ttttaccaga caaccattac 600
ctgtcgacac aatctgccct ttcgaaagat cccaacgaaa agcgtgacca catggtcctt 660
cttgagtttg taactgctgc tgggattaca catggcatgg atgaactata caaataa 717

<210> 8
<211> 717
<212> DNA
<213> Artificial Sequence

<220>

<223> synthetic construct

<400> 8
atgagtaaag gagaagaact tttcactgga gttgtcccaa ttcttgttga attagatggt 60
gatgttaatg ggcacaaatt ttctgtcagt ggagagggtg aaggtgatgc aacatacgga 120
aaacttaccc ttaaatttat ttgcactact ggaaaactac ctgttccatg gccaacactt 180
gtcactactt tctcttatgg tgttcaatgc ttttcaagat acccagatca tatgaaacgg 240
catgactttt tcaagagtgc catgcccgaa ggttatgtac aggaaagaac tatatttttc 300
aaggatgacg ggaactacaa gacacgtgct gaagtcaagt ttgaaggtga tacccttgtt 360
aatagaatcg agttaaaagg tattgatttt aaagaagatg gaaacattct cggacacaaa 420
ctcgagtaca actataactc acacaatgta tacatcatgg cagacaaaca aaagaatgga 480
atcaaagtta acttcaaaat tcgccacaac attgaagatg gatccgttca actagcagac 540
cattatcaac aaaatactcc aattggcgat ggccctgtcc ttttaccaga caaccattac 600
ctgtccacac aatctgccct ttcgaaagat cccaacgaaa agagagacca catggtcctt 660
cttgagtttg taacagctgc tgggattaca catggcatgg atgaactata caaataa 717

<210> 9
<211> 795
<212> DNA
<213> Artificial Sequence

<220>
<223> synthetic construct

<400> 9
atggctctag ttgttaaagg taaggtaaat attaatgagt ttatcgatct gtcaaagtct 60
gagaaacttc tcccgtcgat gttcacgcct gtaaagagtg ttatggtttc aaaggttgat 120
aagattatgg tccatgaaaa tgaatcattg tctgaagtaa atctcttaaa aggtgtaaaa 180
cttatagaag gtgggtatgt ttgcttagtt ggtcttgttg tgtccggtga gtggaattta 240
ccagataatt gccgtggtgg tgtgagtgtc tgcatggttg acaagagaat ggaaagagcg 300
gacgaagcca cactggggtc atattacact gctgctgcta aaaagcggtt tcagtttaaa 360
gtggtcccaa attacggtat tactacaaag gatgcagaaa agaacatatg gcaggtctta 420
gtaaatatta aaaatgtaaa aatgagtgcg ggctactgcc ctttgtcatt agaatttgtg 480
tctgtgtgta ttgtttataa aaataatata aaattgggtt tgagggagaa agtaacgagt 540
gtgaacgatg gaggacccat ggaactttca gaagaagttg ttgatgagtt catggagaat 600
gttccaatgt cggttagact cgcaaagttt cgaaccaaat cctcaaaaag aggtccgaaa 660
aataataata atttaggtaa ggggcgttca ggcggaaggc ctaaaccaaa aagttttgat 720
gaagttgaaa aagagtttga taatttgatt gaagatgaag ccgagacgtc ggtcgcggat 780
tctgattcgt attaa 795

<210> 10
<211> 807
<212> DNA
<213> Artificial Sequence

<220>
<223> synthetic construct

<400> 10
atggctctag ttgttaaagg aaaagtgaat atcaatgagt ttatcgacct gacaaaaatg 60
gagaagatct taccgtcgat gtttacccct gtaaagagtg ttatgtgttc caaagttgat 120
aaaataatgg ttcatgagaa tgagtcattg tcaggggtga accttcttaa aggagttaag 180
cttattgata gtggatacgt ctgtttagcc ggtttggtcg tcacgggcga gtggaacttg 240
cctgacaatt gcagaggagg tgtgagcgtg tgtctggtgg acaaaaggat ggaaagagcc 300
gacgaggcca ctctcggatc ttactacaca gcagctgcaa agaaaagatt tcagttcaag 360

gtcgttccca attatgctat aaccacccag gacgcgatga aaaacgtctg gcaagtttta 420
gttaatatta gaaatgtgaa gatgtcagcg ggtttctgtc cgctttctct ggagtttgtg 480
tcggtgtgta ttgtttatag aaataatata aaattaggtt tgagagagaa gattacaaac 540
gtgagagacg gagggcccat ggaacttaca gaagaagtcg ttgatgagtt catggaagat 600
gtccctatgt cgatcaggct tgcaaagttt cgatctcgaa ccggaaaaaa gagtgatgtc 660
cgcaaaggga aaaatagtag tagtgatcgg tcagtgccga acaagaacta tagaaatgtt 720
aaggattttg gaggaatgag ttttaaaaag aataatttaa tcgatgatga ttcggaggct 780
actgtcgccg aatcggattc gttttaa 807

<210> 11
<211> 795
<212> DNA
<213> Artificial Sequence

<220>
<223> synthetic construct

<400> 11
atggctctag ttgttaaagg taaggtaaat attaatgagt ctatcgatct gtcaaagtct 60
gagaaacttc tcccgtcgat gttcacgcct gtaaagagtg ttatggtttc aaaggttgat 120
aagattatgg tccatgaaaa tgaatcattg tctgaagtaa atctcttaaa aggtgtaaaa 180
cttatagaag gtgggtatgt ttgcttagtt ggtcttgttg tgtccggtga gtggaattta 240
ccagataatt gccgtggtgg tgtgagtgtc tgcatggttg acaagagaat ggaaagagcg 300
gacgaagcca cactggggtc atattacact gctgctgcta aaaagcggtt tcagttcaag 360
gtcgttccca attatgctat aaccacccag gatgcagaaa agaacatatg gcaggtctta 420
gtaaatatta aaaatgtaaa aatgagtgcg ggctactacc ctttgtcatt agaatttgtg 480
tctgtgtgta ttgtttataa aaataatata aaattgggtt tgagggagaa agtaacgagt 540
gtgaacgatg gaggacccat ggaactttca gaagaagttg ttgatgagtt catggagaat 600
gttccaatgt cgatcaggct tgcaaagttt cgaaccaaat cctcaaaaag aggtccgaaa 660
aataataata atttaggtaa ggggcgttca ggcggaaggc ctaaaccaag aagttttgat 720
gaagttgaaa aagagtttga taatttgatt gaagatgaag ccgagacgtc ggtcgcggat 780
tctgattcgt attaa 795

<210> 12
<211> 795
<212> DNA
<213> Artificial Sequence

<220>
<223> synthetic construct

<400> 12
atggctctag ttgttaaagg taaggtaaat attaatgagt ttatcgatct gtcaaagtct 60
gagaaacttc tcccgtcgat gttcacgcct gtaaagagtg ttatggtttc aaaggttgat 120
aagattatgg tccatgaaaa tgaatcattg tctgaagtaa atctcttaaa aggtgtaaaa 180
cttatagaag gtgggtatgt ttgcttagtt ggtcttgttg tgtccggtgt gtggaattta 240
ccagataatt gccgtggtgg tgtgagtgtc tgcatggttg acaagagaat ggaaagagcg 300
gacgaggcca cactcggatc ttactacact gctgctgcta aaaagcggtt tcagttcaag 360
gtcgttccca attatgctat aaccacccag gatgcagaaa agaacatatg gcaggtctta 420
gtaaatatta aaaatgtaaa aatgagtgcg ggctactgcc ctttgtcatt agaatttgtg 480
tctgtgtgta ttgtttataa aaataatata aaattgggtt tgagggagaa agtaacgagt 540
gtgaacgatg gaggacccat ggaactttca gaagaagttg ttgatgagtt catggagaat 600
gttccaatgt cggttagact cgcaaagttt cgaaccaaat cctcaaaaag aggtccgaaa 660
aataataata atttaggtaa ggggcgttca ggcggaaggc ctaaaccaaa aagttttgat 720
gaagttggaa aagagtttga taatttgatt gaagatgaag ccgagacgtc ggtcgcggat 780

tctgattcgt attaa 795

<210> 13
<211> 795
<212> DNA
<213> Artificial Sequence

<220>
<223> synthetic construct

<400> 13
atggctctag ttgttaaagg taaggtaaat attaatgagt ttatcgatct gtcaaagtct 60
gagaaacttc tcccgtcgat gttcacgcct gtaaggagtg ttatggtttc aaaggttgat 120
aagattatgg tccatgaaaa tgaatcattg tctgaagtaa atctcttaaa aggtgtaaaa 180
cttatagaag gtgggtatgt ttgcttagtt ggtcttgttg tgtccggtga gtggaattta 240
ccagataatt gccgtggtgg tgtgagtgtc tgcatggttg acaagagaat ggaaagagcg 300
gacgaagcca cactggggtc atattacact gctgctgcta aaaagcggtt tcagtttaaa 360
gtggtcccaa attacggtat tactacccag gacgcgatga aaaacgtctg gcaggtctta 420
gtaaatatta aaaatgtaaa aatgagtgcg ggctactgcc ctttgtcatt agaatttgtg 480
tctgtgtgta ttgtttataa aaataatata aaattgggtt tgagggagaa agtaacgagt 540
gtgaacgatg gaggacccat ggaactttca gaagaagttg ttgatgagtt catggagaat 600
gttccaatgt cgatcagact cgcaaagttt cgaaccaaat cctcaaaaag aggtccgaaa 660
aataataata atttaggtaa ggggcgttca ggcggaaggc ctaaaccaaa aagttttgat 720
gaagttgaaa aagagtttga taatttgatt gaagatgaag ccgagacgtc ggtcgcggat 780
tctgattcgt attaa 795

<210> 14
<211> 796
<212> DNA
<213> Artificial Sequence

<220>
<223> synthetic construct

<400> 14
atggctctag ttgttaaagg taaggtaaat attaatgagt ttatcgatct gtcaaagtct 60
gagaaacttc tcccgtcgat gttcacgcct gtaaagagtg ttatggtttc aaaggttgat 120
aagattatgg tccatgaaaa tgaatcattg tctgaagtaa atctcttaaa aggtgttaag 180
cttattgata gtggatacgt ctgtttagcc ggtttggtcg tcacgggcga gtggaattta 240
ccagataatt gccgtggtgg tgtgagtgtc tgcatggttg acaagagaat ggaaagagcg 300
gacgaagcca cactggggtc atattacact gctgctgcta aaaagcggtt tcagttcaag 360
gtcgttccca aattacggta ttactaccca ggatgcagaa aagaacatat ggcaggtctt 420
agtaaatatt aaaaatgtaa aaatgagtgc gggctactgc ccgctttctc tggagtttgt 480
gtctgtgtgt attgtttata aaaataatat aaaattgggt ttgagggaga aagtaacgag 540
tgtgaacgat ggaggaccca tggaactttc agaagaagtt gttgatgagt tcatggagaa 600
tgttccaatg tcggttagac tcgcaaagtt tcgaaccaaa tcctcaaaaa gaggtccgaa 660
aaataataat aatttaggta aggggcgttc aggcggaagg cctaaaccaa aaagttttga 720
tgaagttgaa aaagagtttg ataatttgat tgaggatgat tcggaggcta ctgtcgccga 780
ttctgattcg tattaa 796

<210> 15
<211> 795
<212> DNA
<213> Artificial Sequence

<220>
<223> synthetic construct

<400> 15
atggctctag ttgttaaagg aaaagtgaat attaatgagt ttatcgatct gtcaaagtct 60
gagaaacttc tcccgtcgat gttcacgcct gtaaagagtg ttatggtttc aaaggttgat 120
aagattatgg tccatgaaaa tgaatcattg tctgaagtaa atctcttaaa aggtgtaaaa 180
cttatagaag gtgggtatgt ttgcttagtt ggtcttgttg tgtccggcga gtggaattta 240
ccagataatt gccgtggtgg tgtgagtgtc tgcatggttg acaagagaat ggaaagagcg 300
gacgaagcca cactggggtc atattacact gctgctgcaa agaaaagatt tcagttcaag 360
gtcgttccca attatgctat aaccacccag gatgcagaaa agaacatatg gcgggtctta 420
gtaaatatta aaaatgtaaa aatgagtgcg ggctactgcc cgctttctct ggagtttgtg 480
tctgtgtgta ttgtttataa aaataatata aaattgggtt tgagggagaa agtaacgagt 540
gtgaacgatg aaggacccat ggaactttca gaagaagttg ttgatgagtt catggagaat 600
gttccaatgt cgatcaggct cgcaaagttt cgaaccaaat cctcaaaaag aggtccgaaa 660
aataataata atttaggtaa ggggcgttca ggcggaaggc ctaaaccaaa aagttttgat 720
gaagttgaaa aagagtttga taatttgatt gaagatgaag ccgagacgtc ggtcgcggat 780
tctgattcgt actaa 795

【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
10 October 2002 (10.10.2002)

PCT

(10) International Publication Number
WO 02/079468 A3

(51) International Patent Classification[7]: C12N 15/10

(21) International Application Number: PCT/US02/03055

(22) International Filing Date: 1 February 2002 (01.02.2002)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/266,386 2 February 2001 (02.02.2001) US
60/268,785 14 February 2001 (14.02.2001) US

(63) Related by continuation (CON) or continuation-in-part (CIP) to earlier applications:
US 60/266,386 (CIP)
Filed on 2 February 2001 (02.02.2001)
US 60/268,785 (CIP)
Filed on 14 February 2001 (14.02.2001)

(71) Applicant *(for all designated States except US)*: **LARGE SCALE BIOLOGY CORPORATION** [US/US]; 3333 Vaca Valley Parkway, Suite 1000, Vacaville, CA 95688 (US).

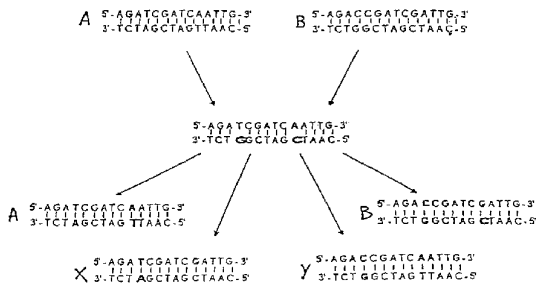
(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: **PADGETT, Hal, S.** [US/US]; 638 Blossom Oak Court, Vacaville, CA 95687 (US). **FITZMAURICE, Wayne, P.** [US/US]; 1218 Las Encinas Court, Vacaville, CA 95688 (US). **LINDO, John, A.** [US/US]; 143 Sundance Drive, Vacaville, CA 95688 (US).

(74) Agents: **WONG, James, J.** et al.; Campbell & Flores LLP, 4370 La jolla Village Drive, 7th Floor, San Diego, CA 92122 (US).

(81) Designated States *(national)*: AE, AG, AL, AM, AT (utility model), AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ (utility model), CZ, DE (utility model), DE, DK (utility model), DK, DM, DZ, EC, EE (utility model), EE, ES, FI (utility model), FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD,

[Continued on next page]

(54) Title: A METHOD OF INCREASING COMPLEMENTARITY IN A HETERODUPLEX POLYNUCLEOTIDE



(57) Abstract: We describe here an *in vitro* method of increasing complementarity in a heteroduplex polynucleotide sequence. The method uses annealing of opposite strands to form a polynucleotide duplex with mismatches. The heteroduplex polynucleotide is combined with an effective amount of enzymes having strand cleavage activity, 3' to 5' exonuclease activity, and polymerase activity, and allowing sufficient time for the percentage of complementarity to be increased within the heteroduplex. Not all heteroduplex polynucleotide will necessarily have all mismatches resolved to complementary. The resulting polynucleotide is optionally ligated. Several variant polynucleotide result. At sites where either of the opposite strands has templated recoding in the other strand, the resulting percent complementarity of the heteroduplex polynucleotide sequence is increased. The parent polynucleotides need not be cleaved into fragments prior to annealing heterologous strands. Therefore, no reassembly is required.

WO 02/079468 A3

WO 02/079468 A3

SE, SG, SI, SK (utility model), SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZM, ZW.

(84) **Designated States** *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, 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:**
27 March 2003

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.

【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
10 October 2002 (10.10.2002)

PCT

(10) International Publication Number
WO 02/079468 A3

(51) International Patent Classification⁷: C12N 15/10

(21) International Application Number: PCT/US02/03055

(22) International Filing Date: 1 February 2002 (01.02.2002)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/266,386 2 February 2001 (02.02.2001) US
60/268,785 14 February 2001 (14.02.2001) US

(63) Related by continuation (CON) or continuation-in-part (CIP) to earlier applications:
US 60/266,386 (CIP)
Filed on 2 February 2001 (02.02.2001)
US 60/268,785 (CIP)
Filed on 14 February 2001 (14.02.2001)

(71) Applicant *(for all designated States except US)*: **LARGE SCALE BIOLOGY CORPORATION** [US/US]; 3333 Vaca Valley Parkway, Suite 1000, Vacaville, CA 95688 (US).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **PADGETT, Hal, S.** [US/US]; 638 Blossom Oak Court, Vacaville, CA 95687 (US). **FITZMAURICE, Wayne, P.** [US/US]; 1218 Las Encinas Court, Vacaville, CA 95688 (US). **LINDO, John, A.** [US/US]; 143 Sundance Drive, Vacaville, CA 95688 (US).

(74) Agents: **WONG, James, J.** et al.; Campbell & Flores LLP, 4370 La jolla Village Drive, 7th Floor, San Diego, CA 92122 (US).

(81) Designated States *(national)*: AE, AG, AL, AM, AT (utility model), AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ (utility model), CZ, DE (utility model), DE, DK (utility model), DK, DM, DZ, EC, EE (utility model), EE, ES, FI (utility model), FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK (utility model), SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZM, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, 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:
27 March 2003

(48) Date of publication of this corrected version:
4 December 2003

(15) Information about Correction:
see PCT Gazette No. 49/2003 of 4 December 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 02/079468 A3

(54) Title: A METHOD OF INCREASING COMPLIMENTARITY IN A HETERODUPLEX POLYNUCLEOTIDE

(57) Abstract: We describe here an *in vitro* method of increasing complementarity in a heteroduplex polynucleotide sequence. The method uses annealing of opposite strands to form a polynucleotide duplex with mismatches. The heteroduplex polynucleotide is combined with an effective amount of enzymes having strand cleavage activity, 3' to 5' exonuclease activity, and polymerase activity, and allowing sufficient time for the percentage of complementarity to be increased within the heteroduplex. Not all heteroduplex polynucleotide will necessarily have all mismatches resolved to complementary. The resulting polynucleotide is optionally ligated. Several variant polynucleotide result. At sites where either of the opposite strands has templated recoding in the other strand, the resulting percent complementarity of the heteroduplex polynucleotide sequence is increased. The parent polynucleotides need not be cleaved into fragments prior to annealing heterologous strands. Therefore, no reassembly is required.

【国際調査報告】

INTERNATIONAL SEARCH REPORT

International Application No
PCT/US 02/03055

A. CLASSIFICATION OF SUBJECT MATTER
IPC 7 C12N15/10

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)
IPC 7 C12N C12Q

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 practical, search terms used)
EPO-Internal, BIOSIS, MEDLINE

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	SOLARO P C ET AL: "Endonuclease VII of phage T4 triggers mismatch correction in vitro" JOURNAL OF MOLECULAR BIOLOGY, LONDON, GB, vol. 230, 1993, pages 868-877, XP002108749 ISSN: 0022-2836 cited in the application page 869, right-hand column, paragraph 4 -page 870, left-hand column, paragraph 1; figures 5,8 page 873, left-hand column, paragraph 2 - paragraph 4 page 875, right-hand column, paragraph 2 - paragraph 4 --- -/--	2-16,20, 22,26

[X] Further documents are listed in the continuation of box C. [X] Patent family members are listed in 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 document but 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	Date of mailing of the international search report
28 November 2002	19/12/2002

Name and mailing address of the ISA
European Patent Office, P.B. 5818 Patentlaan 2
NL - 2280 HV Rijswijk
Tel. (+31-70) 340-2040, Tx. 31 651 epo nl,
Fax: (+31-70) 340-3016

Authorized officer
Rutz, B

Form PCT/ISA/210 (second sheet) (July 1992)

INTERNATIONAL SEARCH REPORT

International Application No
PCT/US 02/03055

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 5 679 522 A (SU SHIN-SAN ET AL) 21 October 1997 (1997-10-21) column 17, line 10-60; example 1; table 3	2,3, 5-12,14, 15
Y	WO 99 29902 A (VOLKOV ALEXANDER;SHAO ZHIXIN ; ARNOLD FRANCES (US); CALIFORNIA INS) 17 June 1999 (1999-06-17) page 17, line 30 -page 18, line 3 page 20, line 24 -page 21, line 2; claim 2	1-65
Y	OLEYKOWSKI C A ET AL: "MUTATION DETECTION USING A NOVEL PLANT ENDONUCLEASE" NUCLEIC ACIDS RESEARCH, OXFORD UNIVERSITY PRESS, SURREY, GB, vol. 26, no. 20, 1998, pages 4597-4602, XP002943289 ISSN: 0305-1048 cited in the application page 4601, right-hand column, paragraph 2 -page 4602, right-hand column, paragraph 2	1-65
A	VOLKOV A A ET AL: "Recombination and chimeragenesis by in vitro heteroduplex formation and in vivo repair." NUCLEIC ACIDS RESEARCH. ENGLAND 15 SEP 1999, vol. 27, no. 18, 15 September 1999 (1999-09-15), page e18 XP002222860 ISSN: 1362-4962 cited in the application page E18, right-hand column, paragraph 1; figure 3	
A	OLEYKOWSKI CATHERINE A ET AL: "Incision at nucleotide insertions/deletions and base pair mismatches by the SP nuclease of spinach." BIOCHEMISTRY, vol. 38, no. 7, 16 February 1999 (1999-02-16), pages 2200-2205, XP002222861 ISSN: 0006-2960 cited in the application page 2203, right-hand column, paragraph 2 -page 2204, right-hand column, paragraph 1	
A	US 5 869 245 A (YEUNG ANTHONY T) 9 February 1999 (1999-02-09) column 4, line 1-22; examples 2,3 column 7, line 29 - line 50 column 8, line 36 - line 44	

-/--

Form PCT/ISA/210 (continuation of second sheet) (July 1992)

INTERNATIONAL SEARCH REPORT

International Application No
PCT/US 02/03055

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication,where appropriate, of the relevant passages	Relevant to claim No.
A	WO 96 40902 A (TREVIGEN INC) 19 December 1996 (1996-12-19) page 7, line 5-20 page 8, line 34 -page 9, line 6; figure 5 page 10, line 1-9; figure 10 page 26, line 22 -page 28, line 35; example 5	
A	YANG BING ET AL: "Purification, cloning, and characterization of the CEL I nuclease." BIOCHEMISTRY, vol. 39, no. 13, 4 April 2000 (2000-04-04), pages 3533-3541, XP002222862 ISSN: 0006-2960 cited in the application	
E	WO 02 24953 A (BERNSTEIN JEFFREY ;MERCK & CO INC (US); MOORE JEFFREY C (US); MCCA) 28 March 2002 (2002-03-28) page 5, line 15 -page 6, line 26; figures 4-8; examples 3-5	2-16,20, 22, 26-38, 42-45, 51-65

Form PCT/ISA/210 (continuation of second sheet) (July 1992)

INTERNATIONAL SEARCH REPORT

Information on patent family members

Int ional Application No
PCT/US 02/03055

Patent document cited in search report		Publication date	Patent family member(s)		Publication date
US 5679522	A	21-10-1997	US	5556750 A	17-09-1996
			AT	198774 T	15-02-2001
			CA	2175490 A1	11-05-1995
			DE	69426608 D1	22-02-2001
			DE	69426608 T2	02-08-2001
			EP	0788555 A1	13-08-1997
			JP	9504437 T	06-05-1997
			WO	9512688 A1	11-05-1995
			US	5861482 A	19-01-1999
			US	5702894 A	30-12-1997
			US	5858754 A	12-01-1999
			US	5459039 A	17-10-1995
			US	6008031 A	28-12-1999
WO 9929902	A	17-06-1999	AU	746786 B2	02-05-2002
			AU	1624199 A	28-06-1999
			CA	2313380 A1	17-06-1999
			EP	1036198 A1	20-09-2000
			JP	2002509697 T	02-04-2002
			WO	9929902 A1	17-06-1999
			US	2002137661 A1	26-09-2002
US 5869245	A	09-02-1999	AU	721493 B2	06-07-2000
			AU	3284597 A	05-01-1998
			EP	0964929 A1	22-12-1999
			JP	2000511774 T	12-09-2000
			WO	9746701 A1	11-12-1997
WO 9640902	A	19-12-1996	US	5656430 A	12-08-1997
			AU	6252796 A	30-12-1996
			EP	0832210 A1	01-04-1998
			JP	11506932 T	22-06-1999
			WO	9640902 A1	19-12-1996
			US	5763178 A	09-06-1998
WO 0224953	A	28-03-2002	WO	0224953 A1	28-03-2002

Form PCT/ISA/210 (patent family annex) (July 1992)

フロントページの続き

(81)指定国 AP(GH,GM,KE,LS,MW,MZ,SD,SL,SZ,TZ,UG,ZM,ZW),EA(AM,AZ,BY,KG,KZ,MD,RU,TJ,TM),EP(AT,BE,CH,CY,DE,DK,ES,FI,FR,GB,GR,IE,IT,LU,MC,NL,PT,SE,TR),OA(BF,BJ,CF,CG,CI,CM,GA,GN,GQ,GW,ML,MR,NE,SN,TD,TG),AE,AG,AL,AM,AT,AU,AZ,BA,BB,BG,BR,BY,BZ,CA,CH,CN,CO,CR,CU,CZ,DE,DK,DM,DZ,EC,EE,ES,FI,GB,GD,GE,GH,GM,HR,HU,ID,IL,IN,IS,JP,KE,KG,KP,KR,KZ,LC,LK,LR,LS,LT,LU,LV,MA,MD,MG,MK,MN,MW,MX,MZ,NO,NZ,OM,PH,PL,PT,RO,RU,SD,SE,SG,SI,SK,SL,TJ,TM,TN,TR,TT,TZ,UA,UG,US,UZ,VN,YU,ZA,ZM,ZW

(74)代理人 100082898
弁理士 西山 雅也
(72)発明者 パジェット, ハル エス.
アメリカ合衆国, カリフォルニア ９５６８７, ベイカビル, ブラッサム オーク コート ６３８
(72)発明者 フィッツモーリス, ウェイン ピー.
アメリカ合衆国, カリフォルニア ９５６８８, ベイカビル, ラス エンチナス コート １２１８
(72)発明者 リンドボー, ジョン エー.
アメリカ合衆国, カリフォルニア ９５６８８, ベイカビル, サンダンス ドライブ １４３

Ｆターム(参考) 4B024 AA20 BA80 CA02 HA01
4B064 AG01 CA21 CC24 DA20