実時間複製反応モニタリングによる小型並列高速ＤＮＡシーケンサーの開発

研究課題番号　16510155

平成 16 度～平成 17 年度科学研究費補助金　基盤研究(C)(2)

研究成果報告書

平成１８年５月

研究代表者　伏　見　　譲

（埼玉大学大学院理工学研究科教授）

はじめに

進化分子工学実験から生ずる多種の変異体の DNA シーケンシングや、SNP 解析のための DNA シーケンシングは　配列がほとんど同じものの少しの違いを多量の試料に対して行う必要がある。この目的に適した DNA シーケンサーを検討した。

DNA シーケンシングのサンガー法は、当初、放射能標識した DNA 複製反応の生産物を、電気泳動で空間展開し、それをオートラジオグラムで可視化し読み取るという多段階のステップを取る遅いプロセスであった。ヒトゲノム計画を推進した DNA シーケンサーは、この電気泳動を時間展開に変え、４色蛍光標識した DNA 複製産物を電気泳動時にレーザーで実時間モニターするという高速化を実現したものであった。さらに高速化するには、DNA 複製反応時に何らかの方法で実時間モニターすればよい。パイロシーケンシング法はその一つの解答であった。われわれはこれとは別に、DNA 複製反応を実時間で電気化学的にモニターすることによって配列を決定する並列高速シーケンサーを提案した。その基本原理は、DNA 複製反応が１塩基伸長に伴い、平均 0.3 個のプロトンが吸収されること（at pH7）に基づく。半導体ｐＨセンサーは微小・高集積・高速応答という特長をもつので、超並列モニターが可能となり、高スループットのシーケンサーとなることが期待できる。欠点としては、現状では短鎖 DNA にしか適用できない。しかし、進化分子工学実験から生ずる多種の変異体の DNA シーケンシングや、SNP 解析のための DNA シーケンシングにはこれで十分である。

次のような段階でその実用化を検討した。（１）固定化タグ配列（ブイＤＮＡ）の固定化密度は従来のコンストラクトでも十分量を達成していたが、今回導入したヘアピン状ブイＤＮＡのヘアピンループ中央に配置したチオ U を介してゲート表面にジスルフィド結合させる方法でも、1.3×10^4 分子/μm^2 という十分量を達成できた。ブイＤＮＡに対する相補タグ配列を持つ未知配列 DNA のハイブリダイゼーションの効率も７０％に達した。この結果、理論検出限界の１０倍量の未知配列ＤＮＡを固定化できたことになる。（２）この試料ＤＮＡをブイＤＮＡから剥離し、ブイＤＮＡを再利用できることを確認した。（３）１対の差動型 ISFET を薄型フローセルに固定化した装置を改良し、セルフプライミング DNA の１０塩基長の伸長に伴うｐＨ変化を観測し、立ち上がり時間 60s で 8.5mV の出力を得た。この出力は１塩基伸長を観測できる大きさである。立ち上がり時間の遅さは、本方法の特長である高速性を損なう。これは表面複製反応の遅さではなく、Microfludics 技術の未熟さに起因する液間界面の平坦化に起因する。

研究組織

研究代表者： 伏見 譲 （埼玉大学大学院理工学研究科教授）
研究協力者： 飯塚 毅 （埼玉大学大学院理工学研究科博士前期課程）

謝辞：ISFET を供与いただいた新電元工業株式会社に謝意を表します。また、先行研究を遂行してくださった研究室メンバー各位に謝意を表します。

交付決定額（配分額）

（金額単位：千円）

	直接経費	間接経費	合 計
平成１６年度	２，１００	０	２，１００
平成１７年度	１，５００	０	１，５００
総 計	３，６００	０	３，６００

研究発表

本研究に広い意味で関連しているものを列挙した。太字の論文はそのコピーを本報告書後半に添付した。

（１）学会誌等

Aita T., Morinaga M., Husimi Y., Thermodynamical Interpretation of Evolutionary Dynamicson a Fitness Landscape in an Evolution Reactor, I, *Bull.Math.Biol.*,**66**, 1371-1403 (2004)

Suzuki M., Ito Y., Savage H.E., Husimi, Y., Douglas, K. T. Protease-sensitive signalling by chemically engineered intramolecular fluorescent resonance energy tranfer (FRET) mutants of Green Fluorescent Protein, *Biochem.Biophys.Acta*, **1679**, 222-229 (2004)

Suzuki, M. Ito, Y., Savage, H., Husimi, Y., Douglas, K T., Introduction of a Protease Indicator Using Intramolecular FETS between GFP and Chemical Compounds into Cells, *The Science and Engineering Reports of Saitama University,* **36**(29), (2004)

Tabuchi I., Soramoto S., Ueno S., Husimi Y., Multi-Line Split DNA Synthesis: a combinatorial method to make a high quality peptide library, *BMC Biotechnol.*,4:9 (2004)

Aita T., Morinaga S., Husimi Y., Thermodynamical Interpretation of Evolutionary Dynamics on a Fitness Landscape in an Evolution Reactor, II, *Bull.Math.Biol.*, 67 (3), 613-661 (2005)

Hamamatsu N., Aita T., Nomiya Y., Uchiyama H., Nakajima M., Husimi Y., Shi banaka Y., Biased mutation-assembling: an efficient method for rapid directed ev olution through simultaneous mutation accumulation, *Protein Eng.Design & Selec tion,* 18(6), 265-271 (2005)

伏見 譲, 試験管内自然淘汰型進化リアクタ, *人工知能学会誌*, 19, 672-677 (2004)

伏見 譲, 試験管内での分子の自己増殖と進化, *パリティ*, 20, 10, 28-29 (2005)

伏見 譲, 偶然から必然へ　分子進化の偶然と必然, *数理科学*, 44, 1, 36-40 (2006)

（2）口頭発表

伏見 譲, 埼玉バイオプロジェクト, *将来技術研究会（ATOM）*（埼玉大学）,2004.6

伏見 譲, in vitro 自然淘汰型進化リアクターにおける進化能の進化, *国際高等研究所「ダイナミックスから見た生命的システムの進化と意義－構造と機能の固定化と変容に関する数理的観点」第２回研究会*（札幌アスペンホテル）, 2004.8

Aita T., Husimi, Y., Thermodynamical Interpretation of Evolution on a Fitness Landscape for a Biopolymer, *5th International Conference on Biological Physics* (Gothenburg), 2004.8

Husimi, Y., Evolution Reactor Dynamics of a Transcripatse-in vitro virus, *5th International Conference on Biological Physics* (Gothenburg), 2004.8

伏見 譲, 進化バイオテクノロジーの動向, *埼玉バイオシンポジウム～進化バイオテクノロジ*

一～, 2004.11

上野 真吾, 深石 圭, 新井 秀直, 鈴木 美穂, 伏見 譲, N 末端側にリンカーを持つ in vitro virus の開発, *日本生物物理学会第 42 回年会講演予稿集,* Vol.44, S215(2004.12)

Biyani Manish, 伏見 譲, 根本 直人, Protein synthesis on bead using in vitro virus: Can immobilized mRNA enhance the translation and folding of protein?, *日本生物物理学会第 42 回年会講演予稿集,* Vol.44, S215(2004.12)

相田 拓洋, 伏見 譲, 実験室内分子進化ダイナミクスの熱力学的及び情報論的概念による解釈, *日本生物物理学会第 42 回年会講演予稿集,* Vol.44, S252(2004.12)

永安 弘樹, 馬 暁涛, 上野 泰生, 伏見 譲, 自然淘汰型核酸進化実験系におけるプロモータの進化, *日本生物物理学会第 42 回年会講演予稿集,* Vol.44, S253(2004.12)

伏見 譲, ～高速分子進化による高機能バイオ分子の創出～研究成果の概要, *埼玉県地域結集型共同研究事業（埼玉バイオ）平成 16 年度研究成果発表会 ～高速分子進化による高機能バイオ分子の創出～*, 2005.3

鈴木美穂, 沖保彦, 伏見譲, 量子ドット上に蛍光ラベルした DNA を固定化する事で観察される蛍光共鳴エネルギー移動(FRET)を利用した DNase 活性感受型バイオプローブについて, *日本分析化学会年会講演要旨集,* 54th, p. 129, 2005.8.

二木類, 鈴木武尊, 伏見譲, 相田拓洋, NADPH を補酵素とする酵素の高速分子進化においてスクリーニングを効率的に行うための酵素活性測定法の改良, *日本分子生物学会年会プログラム・講演要旨集*, 27th, p. 845, 2005.11.

鈴木美穂, 伏見譲, 伊藤洋一郎, DOUGLAS T K., GFP と化学蛍光試薬の分子内蛍光共鳴エネルギー移動(FRET)を用いた DNase 活性感受性バイオプローブについて, *日本分子生物学会年会プログラム・講演要旨集,* 27th, p.897, 2005.11.

沖保彦, 伏見譲, 鈴木美穂, 伊藤洋一郎, DOUGLAS K T, GFP と化学蛍光試薬の分子内 FRET を用いた DNase 活性感受型バイオプローブについて, *日本分子生物学会年会プログラム・講演要旨集,* 27th, p.897, 2005.11.

飯塚毅, 富田允, 田中寿幸, 金村隆広, 伏見譲, 複製反応の実時間電気化学モニターによる

DNA シーケンサの検討, *日本生物物理学会第 43 回年会講演予稿集,* Vol.45, S202(2005.11)

新井秀直, 上野真吾, 鈴木美穂, 伏見譲, 新規 in vitro virus の開発, *日本生物物理学会第 43 回年会講演予稿集,* Vol.45, S222(2005.11)

馬暁涛, 永安弘樹, 上野泰生, 国井宇雄, 伏見譲, 等温核酸増幅を用いた進化実験における中立経路を経た進化, *日本生物物理学会第 43 回年会講演予稿集,* Vol.45, S277(2005.11)

相田拓洋, 林勇樹, 豊田一志, 卜部格, 伏見譲, 人工進化実験からの適応度地形特性の抽出:ファージ感染能の進化実験への適用, *日本生物物理学会第 43 回年会講演予稿集,* Vol.45, S278(2005.11)

（3）出版物

伏見 譲, 生命の起源と進化への生物物理学的アプローチ, *学術の動向,* **9**(6), 70-73, (2004)

伏見 譲編,『生命の起源—「物質の進化」から「生命の進化」へ—』(丸善、2004)

猪飼 篤, 中西 守, 桜井 実, 佐藤 衛, 伏見 譲訳.『生命科学のための物理化学』(Tinoco 他著) (東京化学同人、2004)

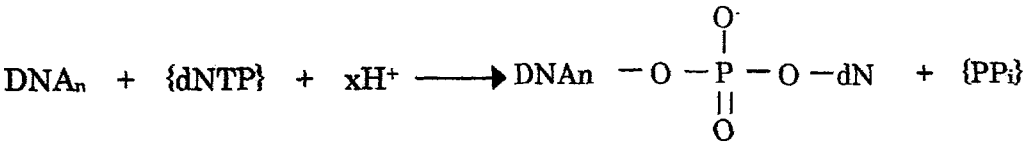
伏見譲, 岩波講座・物理の世界『DNA と遺伝情報の物理』, (岩波書店, 2005)

研究成果

１．原理

１．１　ＤＮＡ複製反応に伴うｐＨ変化

ＤＮＡ複製反応において、１ヌクレオチドの伸長反応は次の式によって表される。



ここで、{・・・} は、プロトン解離状態、Mg^{2+}結合状態の異なる分子種のセットを意味し、各分子種間で異なる pK_a 値をもつ。反応式中の H^+ の係数 x は、溶液の初期ｐＨに依存するが、{dNTP} と {PP_i} の各分子種のｐK_a値から、ｐH7.0 近傍では、約+0.3 であることが計算された。すなわち、試料溶液のｐＨ緩衝能を落としておけば、ｐＨ はアルカリ側にシフトする。この事実は先行研究により ISFETｐＨセンサーの出力電圧として検出され確認された（Sakurai,T. & Husimi,Y., Anal.Chem.64, 1996 (1992)）。

１．２　複製反応の実時間モニターによる超並列ＤＮＡシーケンサーのプロトコール

DNA 複製反応を実時間モニターすることによる高速 DNA シーケンシングの方法を図１に提案する。

チューブの内面に直線上に M 対の ISFET が集積されており、それらのゲート膜表面の窒化ケイ素にプローブ（または、プイと呼ぶ）ssDNA の 3'末端側を固定化する。M＋１種のプローブ ssDNA が固定化されている。M 種の配列未知の試料 ssDNA の 5'末端側にこのプローブのタグ配列に相補な配列を付け、3'末端側にはセルフプライミングするユニバーサル配列を付けておく。この他１種の参照用 ssDNA を用意する。

① M 種未知試料 DNA と参照用 DNA をプローブにハイブリダイズさせる。
② チューブに DNA ポリメラーゼ（KF）とｄATP に流し、次に、洗浄液を流す。
③ KF と dTTP を流し、次に洗浄する。
④ KF と dGTP を流し、次に洗浄する。
⑤ KF と dCTP を流し、次に洗浄する。
⑥ 上記の②－④を未知配列の塩基数だけ繰り返す。
⑦ 変性バッファーを流し、試料 DNA を解離流出させる。
⑧ 上記①－⑦を繰り返す。

②－⑤のいずれかのステップで、DNA 複製反応が１塩基伸長するが、その時、ｐＨがアルカリ側にパルス状に変化する。試料で同じ塩基がｎ個並んでいる場合は、パルスの高さが

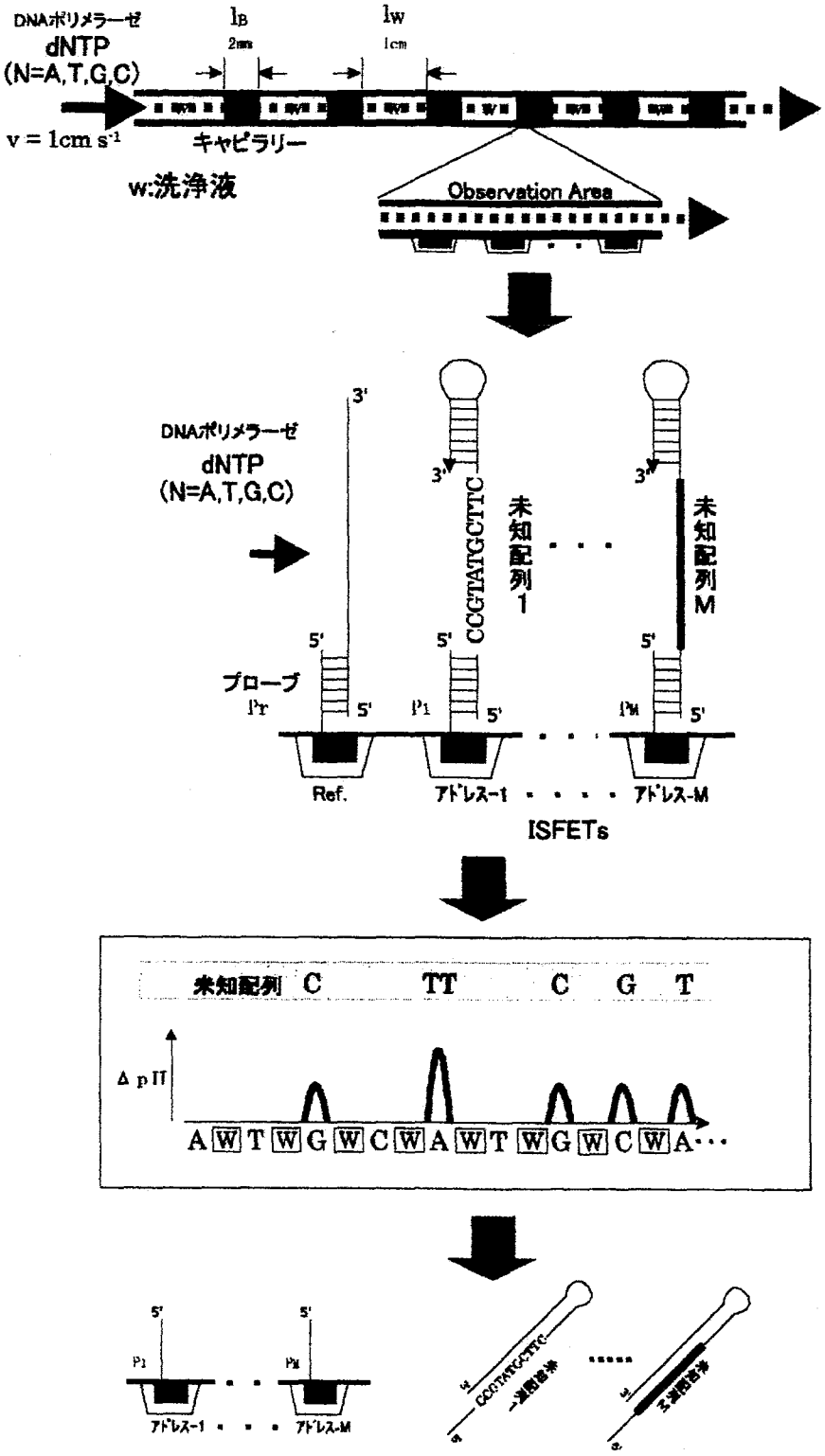


図１．超並列高速 DNA シーケンシングのプロトコール概念図

n倍になることが期待される。実際、M.Ronaghi らのパイロシーケンシング法では、4 塩基の連を識別している。

この方法で1時間に決定できる塩基数は、3600 x M x v / 4(l_W+l_B)。
M=256 の時、250,000 base/hr という高速性が期待できる。

1．4　理論検出限界

ここで用いている ISFET pH センサーの感度は、中性 pH で0．02 pH である。
次に、複製反応に伴う局所的プロトン濃度変化を計算する。
まず、図2に示す、1塩基伸長時のプロトンの拡散距離 L を評価する。プロトンの拡散係数は $D = 9.31 \times 10^{-5}$ cm²/s。固定化面積が 300 μ m² 以上あれば、1次元拡散と見なしてよく、DNA ポリメラーゼの反応速度を u[nt/s]とすると、

$$L = \sqrt{2Du^{-1}} \times 10^{-2} \quad \mathrm{m}$$

DNA の固定化密度を　N 分子/μ m²　とすると、図2の直方体の体積は

$$v = L / N \times 10^{-9} \quad \ell$$

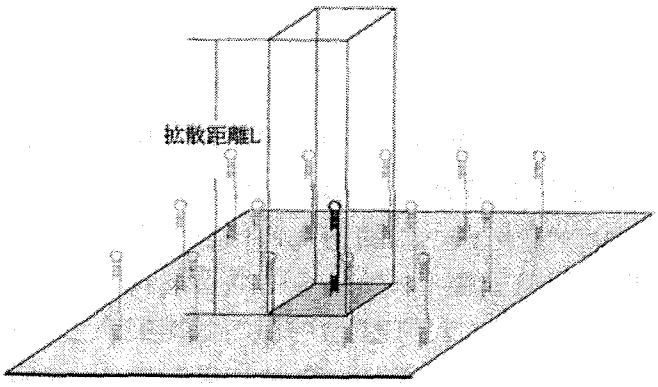


図2．DNA 1分子あたりの作業空間

バッファーの緩衝能を f、アボガドロ数を N_A とすると、体積要素 v における局所 pH 変化パルス高は、pH7.0 において、

$$\Delta pH = -\log\left(10^{-7.0} - 0.3v^{-1} / N_A / f\right) + \log\left(10^{-7.0}\right)$$
$$= -\log\left(10^{-7.0} - \frac{0.3N \times 10^{11}}{N_A f}\sqrt{\frac{u}{2D}}\right) + \log\left(10^{-7.0}\right)$$

これをテーラー展開し、第1項までを取ると、理論検出限界の固定化 DNA 密度は、

$$N = \frac{\Delta pH \times N_A f \ln 10 \times 10^{-7.0}}{0.3 \times 10^{11}} \sqrt{\frac{2D}{u}}$$

ISFET 感度 ΔpH＝0.02、u=25 nt/s、f=5　の時、N=4.0 x 10^3 μ m^{-2}　と評価できる。

DNA 複製反応 1 塩基伸長の反応収率を η とすると、シーケンシング可能な塩基長は、

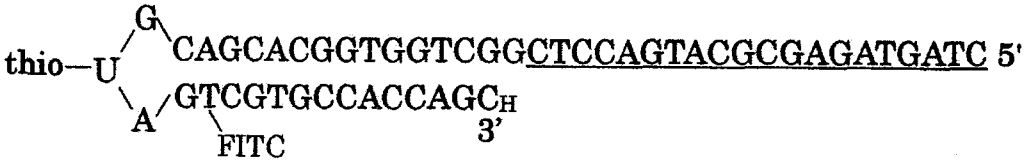
$$L_0 = \log 0.5 / \log \eta$$

η ＝0.9 のとき、Lo=7,　η ＝0.99 のとき、Lo=70。実際、固定化 DNA パイロシーケンシングの場合、15 塩基まで正確に配列決定している。

2．材料と方法

2．1　DNAオリゴマー

プローブssDNA（繋留用DNA、ブイDNAと呼ぶ）としては、次のDNAオリゴマーを合成し用いた。3'末端はジデオキシ体となっており、再利用可能なように伸長反応を抑制している。チオUを介するSS結合でISFETゲート表面に固定化される。



テスト用試料DNAとしては、次のDNAオリゴマーとその誘導体を合成して用いた。アンダーラインの部位で、ブイDNAにハイブリダイズしISFETゲート表面に固定化される。セルフプライミングで複製反応が起こる。誘導体としては、配列決定部分の配列が、

AGTCTGTGGTTTTTTTTT

となっているものであり、dATPの投入により１０塩基、さらにdCTPの投入により、5塩基、さらにｄGTPの投入により3塩基、最後にｄTTPの投入により１塩基伸長することが観察されるような試料である。

2．2　DNAオリゴマーの固定化

チオUを有するブイDNAをスライドガラスやISFETのゲート表面に固定化するには、図３に示すようなSS結合交換反応を用いた。

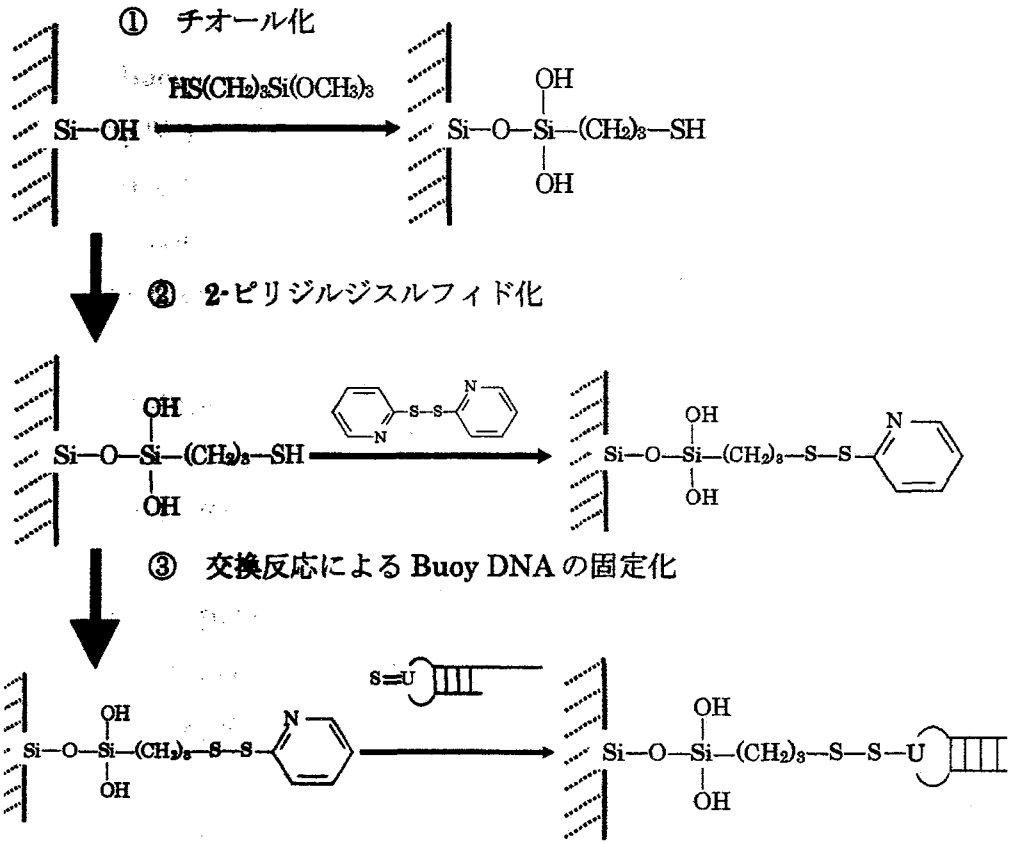


図３　ブイ DNA の固定化反応

２．３　装置

２本の ISFET（新電元工業株式会社から贈与）をオゾンクリーナーで洗浄後、超音波洗浄した。そのうち１本のゲート表面にブイ DNA を固定化し、次いで試料 DNA をハイブリダイズさせた。もう１本の ISFET はシラン化処理を行い REFET とした。

図４のように、PDMA を用いて、２本の ISFET が向かい合わせに固定化されたフローセルを作製した。

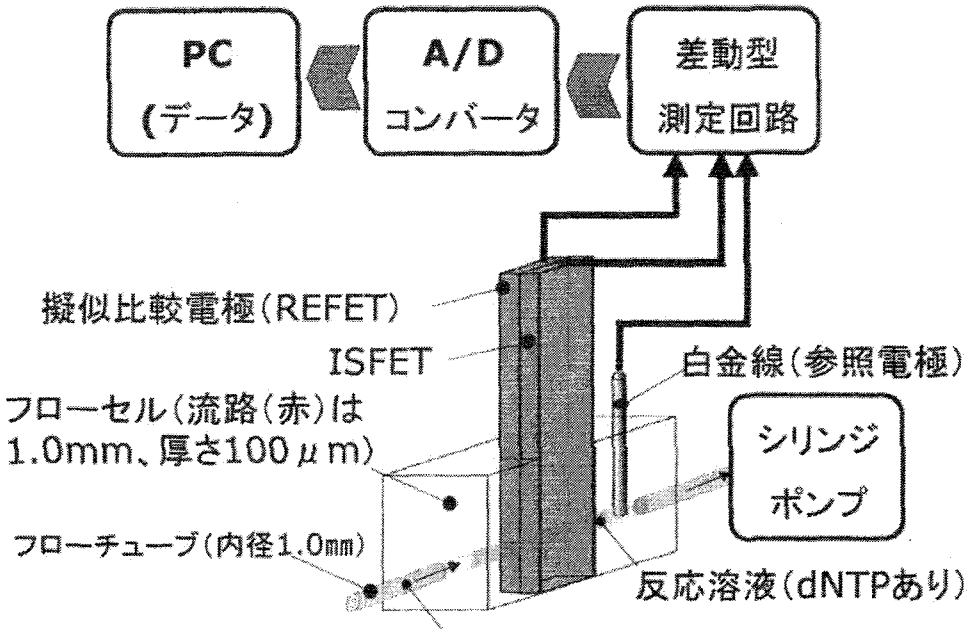


図４．測定装置概略図。差動型測定回路とは、BAS 社製 ISFETmV/pH メーターである。

３．結果

３．１．ＤＮＡ固定化密度

固定化タグ配列（プイＤＮＡ）の固定化密度は従来のコンストラクトでも十分量を達成していたが、今回導入したヘアピン状プイＤＮＡのヘアピンループ中央に配置したチオ U を介してゲート表面にジスルフィド結合させる方法でも、1.3×10^4 分子/μm^2 という十分量を達成できた。図５参照。

プイＤＮＡに対する相補タグ配列を持つ未知配列 DNA のハイブリダイゼーションの効率も７０％に達した。この結果、上記理論検出限界の１０倍量の未知配列ＤＮＡを固定化できたことになる。

この試料ＤＮＡをプイＤＮＡから SDS バッファーにより解離させ、プイＤＮＡを再利用できることを確認した。

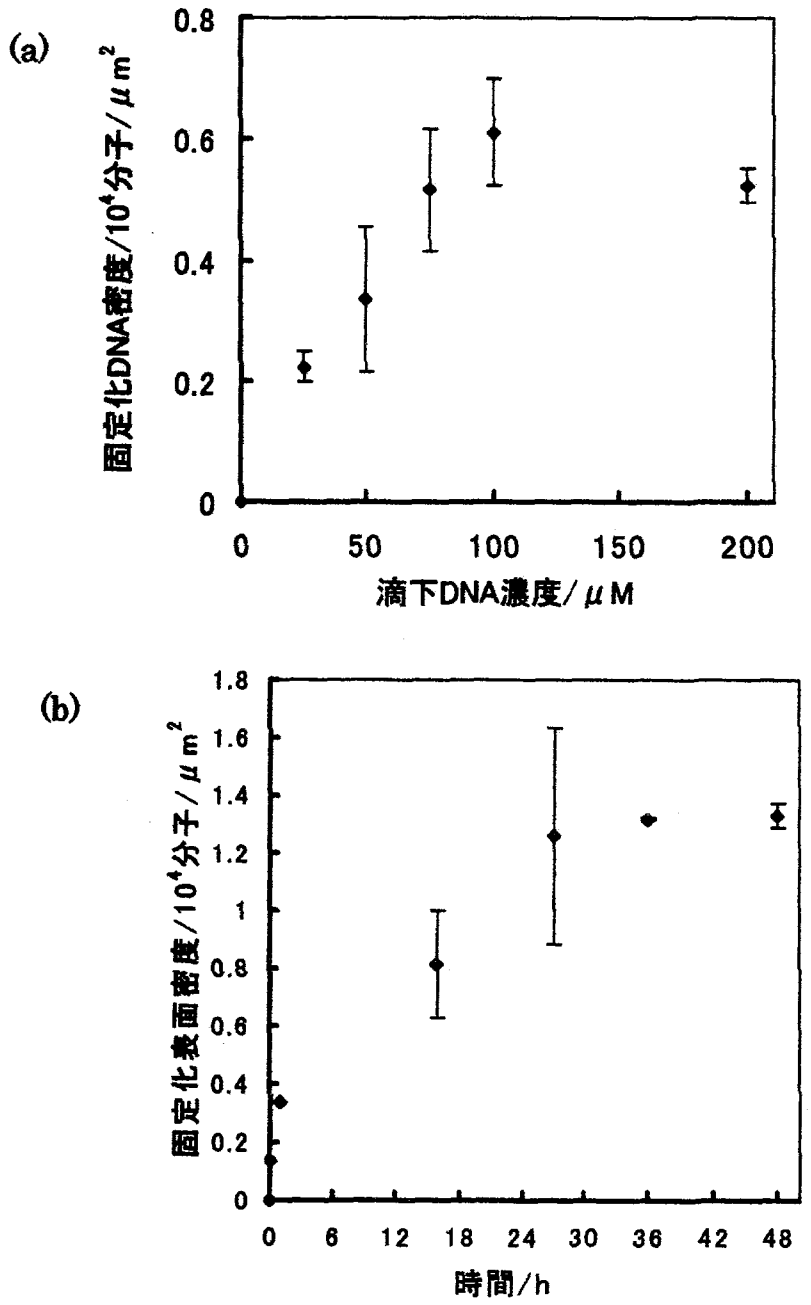


図５．スライドグラスへのブイ DNA の固定化反応における固定化密度。(a) 滴下 DNA 濃度依存性、(b)タイムコース。固定化密度は、FITC 蛍光を蛍光分光光度計で測定し、検量線から表面密度を求めた。

３．２．電気信号出力

図６に示すように、セルフプライミング DNA の１０塩基長の伸長に伴う pH 変化を観測し、90%立ち上がり時間 60s でピーク値 8.5mV の出力を得た。伸長反応が起こらない dTTP のみ

の投入では電気信号出力が無いことを確認してあり、観測された出力は複製反応をモニターしていることになる。この出力は１塩基伸長を観測できる大きさである。

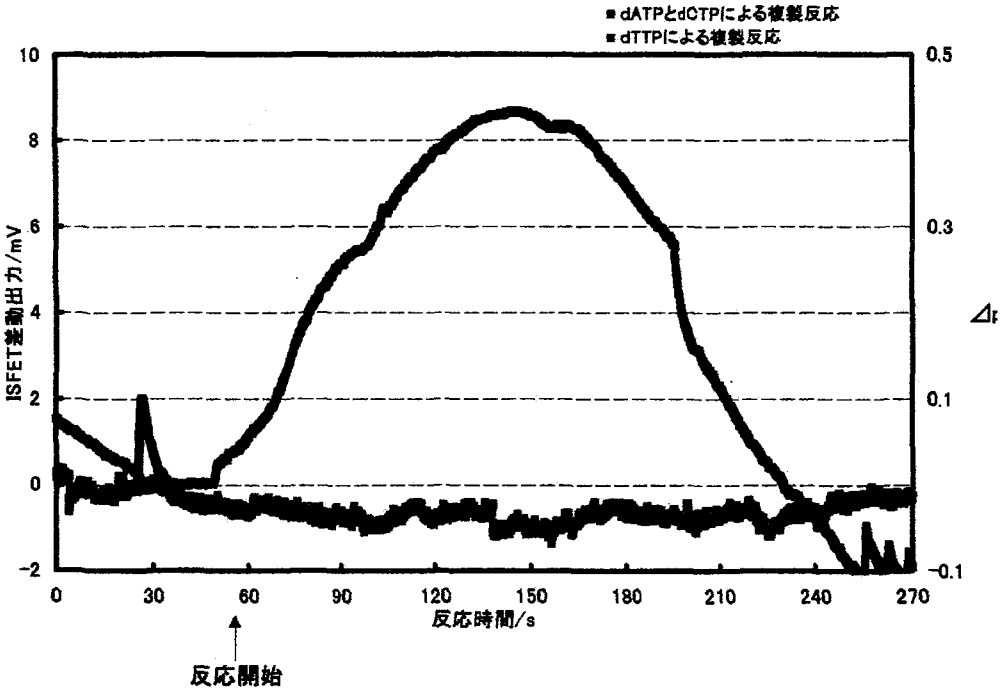


図６．固定化 DNA１０塩基長複製反応の実時間モニター。
右側縦軸はアルカリ方向ｐＨ変化を表す。

試料の塩基配列から考えて伸長反応が起こるはずのないｄTTP のみを投入したときは出力変化は無いが、10 塩基分の伸長反応が起こるｄATP+dCTP を投入すると 8.5mV の出力変化が得られた。

４．今後の展望

図６に示す出力の立ち上がり時間の遅さは、本方法の特長である高速性を損なうことになる。ΔｐＨ＝１の２種の 0.6SSC バッファーを切り替えて流すことによる出力変化曲線の立ち上がりも同程度に遅いことから、この立ち上がりの遅さは表面複製反応の遅さではなく、Microfludics 技術の未熟さに起因する液間界面の平坦化に起因することが判明した。当初の計算通りの高速性を実現するためには、マイクロフローセルの設計を再検討する必要がある。特に ISFET の製造工程と流路製造工程を一体化することが望ましい。

公表した参考論文のコピー集

本シーケンサーの対象である、進化分子工学の関係論文を４編添付して、参考に供する。

BMC Biotechnology

BioMed Central

Methodology article

Open Access

Multi-line split DNA synthesis: a novel combinatorial method to make high quality peptide libraries

Ichiro Tabuchi[1,2], Sayaka Soramoto[2], Shingo Ueno[2] and Yuzuru Husimi*[2]

Address: [1]Tokyo Evolution Research Center, 1-1-45-504, Okubo, Shinjuku-ku, Tokyo 169-0072, Japan and [2]Department of Functional Materials Science, Saitama University,255 Shimo-Okubo, Saitama 338-8570, Japan

Email: Ichiro Tabuchi - tabuchii@mx16.freecom.ne.jp; Sayaka Soramoto - soramoto@evolve.fms.saitama-u.ac.jp; Shingo Ueno - ueshin@evolve.fms.saitama-u.ac.jp; Yuzuru Husimi* - husimi@fms.saitama-u.ac.jp

* Corresponding author

Published: 01 September 2004

BMC Biotechnology 2004, 4:19 doi:10.1186/1472-6750-4-19

Received: 31 March 2004

Accepted: 01 September 2004

This article is available from: http://www.biomedcentral.com/1472-6750/4/19

Abstract

Background: We developed a method to make a various high quality random peptide libraries for evolutionary protein engineering based on a combinatorial DNA synthesis.

Results: A split synthesis in codon units was performed with mixtures of bases optimally designed by using a Genetic Algorithm program. It required only standard DNA synthetic reagents and standard DNA synthesizers in three lines. This multi-line split DNA synthesis (MLSDS) is simply realized by adding a mix-and-split process to normal DNA synthesis protocol. Superiority of MLSDS method over other methods was shown. We demonstrated the synthesis of oligonucleotide libraries with 10^{16} diversity, and the construction of a library with random sequence coding 120 amino acids containing few stop codons.

Conclusions: Owing to the flexibility of the MLSDS method, it will be able to design various "rational" libraries by using bioinformatics databases.

Background

The combinatorial synthesis method has been demonstrating its effectiveness in discovering novel functional molecules. Examples of this method in the field of evolutionary protein engineering are selections of a novel functional peptide from a random library on solid support [1], phage display [2] or *in vitro* virus (synonym for RNA-peptide fusion or mRNA-display) [3-5]. The efficiency of the methods depends on the screening technique employed and the library quality. In the display methods, a library of polynucleotide templates must be prepared in order to obtain a random peptide library. A primitive random library of such templates is $(NNN)_n$ (N = equimolar mixture of A, T, G and C). This library leads to premature short peptides and a particular bias of the amino acid composition, which makes the effective searchable sequence space biased. A slightly improved library NNK or NNS (K / S = equimolar mixture of T and C / G and C) has been conventionally used. Several methods have been developed for a more improved library. Various "rational" libraries in which the nucleotide mixtures were optimized for a target amino acid composition by using a computer calculation have been developed [6-8].

Removal of stop codons to obtain long ORFs is important for the evolutionary design of a novel protein starting from a random library. Several methods based on random block-ligation were reported [9,10]. Two high quality libraries that lead to the successful evolutionary protein design were as follows: the trinucleotide

phosphoramidites (3NPs) method using twenty pre-synthesized trimers of nucleotide phosphoramidites [11-14], and the pre-selecting method using an mRNA display with a C-terminus affinity tag in order to remove stop codons [15].

We report in this article on a convenient method for the construction of a high quality library based on combinatorial DNA synthesis. This library has few stop codon and has an optimized amino acid composition for various purposes. A random library based on the split synthesis [1] is made routinely in combinatorial chemistry, but a few methods [16,17] and a few applications [18,19] have been reported for oligonucleotides synthesis. They were used for mutagenesis and the products did not have high quality for evolutionary protein engineering. We applied the split synthesis to oligodeoxyribonucleotide synthesis and developed a new procedure, based on the synthesis of designed codon mixtures using multi-line DNA synthesizers. Our method, Multi-Line Split DNA Synthesis (MLSDS), requires only standard reagents and three or four synthesizers for DNA synthesis. MLSDS can make various "rational" libraries of huge diversity with few stop codons.

Results and Discussions

Adaptive design to the target amino acid composition

Scheme of the MLSDS method is shown in Fig. 1 and Table 1, and described in detail in Methods section. MLSDS is able to remove not only stop codons but also particular codons. It is able to design the codon composition. We incorporated the effect of the single nucleotide deletion during a general oligonucleotide synthesis [20] into the design.

Designed biased libraries are useful for creating various novel proteins such as a functional peptide without Cys [21] or an engineered protein without Met [22]. Unnatural codons and unnatural amino acid [23] will be also incorporated in desired composition. It will be able to incorporate various results of analysis of bioinformatics databases in order to make an initial library with higher evolvability in experimental protein evolution. The optimum amino acid composition in the library may be different for each target protein. For example, when we want to explore the global protein sequence space exhaustively, the uniform amino acid composition may be the best. When we want to explore only a proven region in the protein sequence space, the use of the average amino acid composition among natural proteins [24] might be better for many aspects. When we want to design a protein with some specific properties, a library with increased or decreased fraction of specific amino acid should be constructed for each segmental region on the polypeptide chain.

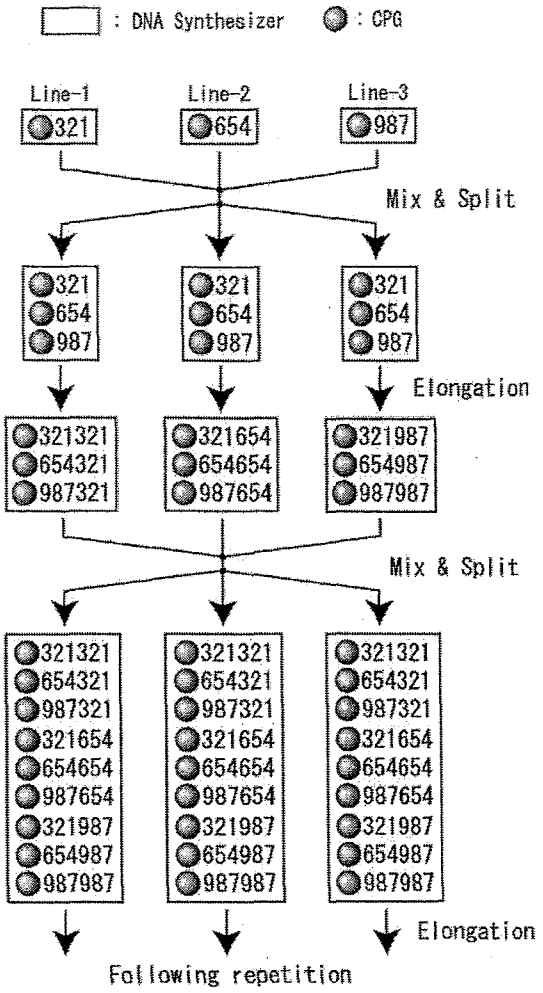


Figure 1

Scheme of the MLSDS method. The case of three-lines is shown. Uppermost three boxes indicate the state in the DNA synthesizers after the first three synthetic cycles, that is, partial mixtures of triplet codons attached to the CPGs. **1**, **2** and **3** denote 1st, 2nd and 3rd letter mixture in the line-1 DNA synthesizer, respectively. Their A:T:G:C mixing ratios are designed with the GA program so that **123** gives a partial mixture of triplet codons without any stop codons. In the same way, **456** and **789** denote corresponding partial mixtures of triplet codons in the line-2 and the line-3 DNA synthesizers, respectively. The equimolar mixture of **123**, **456** and **789**, which is obtained after split-and-mix procedure, can be designed to give approximately the target amino acid composition without stop codons. Examples of **1,2,.....,9** are shown in Table 1.

Table 1: Examples of molar mixing ratio of bases for MLSDS libraries

Library name	No Cys									Natural								
Line Number of Synthesizers	Line-1			Line-2			Line-3			Line-1			Line-2			Line-3		
ID Number of premix phosphoramidites Molar mixing ratio of A:T:G:C	1	2	3	4	5	6	7	8	9	1	2	3	4	5	6	7	8	9
A	2	3	5	5	0	0	2	5	0	5	4	2	2	0	0	3	4	0
T	0	0	3	4	3	0	3	5	2	0	1	1	3	2	3	4	2	5
G	3	0	0	4	3	1	1	0	0	5	0	2	5	5	2	4	0	0
C	3	1	0	0	1	0	1	3	5	3	1	0	2	2	1	2	4	0

Among these wide spectra of requirements, we designed DNA libraries that code peptide libraries having various characteristics and have no stop codons. Examples are: a library with the average amino acid composition of natural proteins [24], which is named "Natural" library in this article, the uniform amino acid composition; and the uniform composition except [Cys] = 0. A library encoding only four kinds of amino acid (a c-Fos mutant library [26]) was also designed. Designed molar mixing ratios of A:T:G:C for some of these libraries are shown in Table 1. Another interesting example was obtained when the target composition was "Uniform except [Met] = 0 and [Term] = 0". The designed molar mixing ratio of A:T:G:C gave the high fitness *F* value (0.96 on three lines splits) and gave no stop codon even if the effect of a point deletion was included in the GA calculation. A Met-less random library may be the best starting library for global search of the protein sequence space. This speculation is supported by the report [22] stating that a mutant dihydrofolate reductase generated by the replacement of all Met had much higher enzymatic activity than the wild type.

Internal deletion problem in the oligonucleotides synthesis process is important. It destroys the codon-based design, leading to stop-codon generation and undesirable amino acid composition. Our program incorporated deletion effects into the GA calculation and succeeded to minimize the deletion problem. Moreover it was reported that contamination of deletion products could be decreased on a denaturing PAGE for DNA of this length [15].

We also investigated the practical number of DNA synthesizers. For this purpose, we calculated the final correlation coefficient between the designed and the various target compositions with up to 6-line DNA synthesizers. As shown in Fig. 2, the final correlation coefficient (= the final fitness) became saturated at about 3- or 4-lines on this program. Our GA program is not the best for obtain best *F* value but suitable for designing actual synthesizing operations. These results showed MLSDS method gave a high quality library even with three DNA synthesizers.

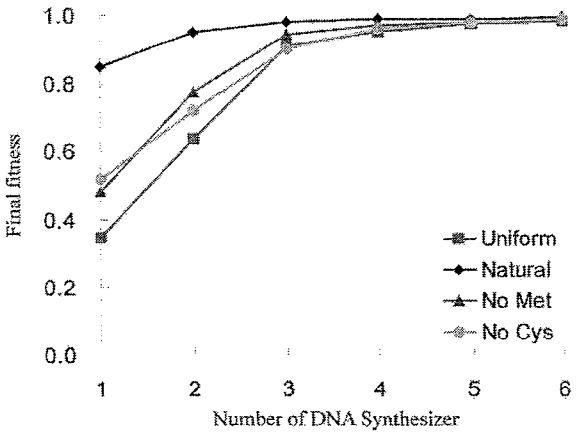


Figure 2
Dependence of the final fitness on the number of DNA synthesizers for various libraries. Ordinate: final fitness (=correlation coefficient). Abscissa: line number of DNA synthesizers. Libraries are "Uniform", "Natural", "No Met" and "No Cys".

When we took the natural abundance as the target amino acid composition, we got a highest fitness value *F* = 0.99 (on three lines) in the GA calculations. This is reasonable, because the average amino acid composition among natural proteins highly correlates to the number of synonymous codons in the standard genetic code table [25].

Synthesis of MLSDS libraries

We synthesized a "Natural" library and a "Uniform except [Cys] = 0" library mentioned in the previous section. In Table 2 the compositions of the actually synthesized DNA libraries are listed in comparison with the target compositions. They were high quality libraries (*F* = 0.85 and 0.66, respectively) without stop codons in full-length DNAs. The deletion rate was about 0.3% per coupling. For the total DNAs including deletants, *F* = 0.90 and 0.60, respectively.

Table 2: Comparison of the amino acid compositions of the actually synthesized DNA libraries (the full length library) with the target composition

Library name	Natural		No Cys		c-Fos e', g'-	
	Target	Actual	Target	Actual	Target	Actual
Ala	8.9%	11.4%	5.3%	6.7%	0.0%	0.0%
Arg	4.7	3.5	5.3	5.9	30.0	29.7
Asn	4.4	3.5	5.3	4.3	0.0	0.0
Asp	5.5	8.6	5.3	5.9	0.0	0.0
Cys	2.8	2.5	**0.0**	**0.0**	0.0	0.0
Gln	3.9	2.1	5.3	5.5	20.0	19.6
Glu	6.2	6.9	5.3	5.5	20.0	25.0
Gly	7.8	12.0	5.3	8.2	0.0	0.0
His	2.0	1.2	5.3	2.7	0.0	0.0
Ile	4.6	3.9	5.3	5.9	0.0	0.0
Leu	7.5	3.7	5.3	4.3	0.0	0.0
Lys	7.0	6.2	5.3	4.7	30.0	27.9
Met	1.7	1.1	5.3	3.5	0.0	0.0
Phe	3.5	4.0	5.3	5.1	0.0	0.0
Pro	4.6	2.5	5.3	2.4	0.0	0.0
Ser	7.1	8.1	5.3	1.6	0.0	0.0
Thr	6.0	5.8	5.3	6.7	0.0	0.0
Trp	1.1	1.8	5.3	8.2	0.0	0.0
Tyr	3.5	4.4	5.3	4.7	0.0	0.0
Val	6.9	6.9	5.3	8.2	0.0	0.0
Term	**0.0**	**0.0**	**0.0**	**0.0**	**0.0**	**0.0**
Number of sequenced codon	-	568	-	255	-	219
Correlation coefficient	-	0.85	-	0.66	-	1.00
Sum of absolute errors	-	26.9%	-	28.1%	-	5.66%

"Term" denotes stop codons. In each target composition, [Term] = 0. "Natural": the library with the natural abundance (the average composition in 207 natural proteins) [18]; "No Cys": the library with uniform composition except [Cys] = 0; "c-Fos e', g'-": a c-Fos mutant library containing only 4 kinds of amino acid as same as in Ref.[26].

We also synthesized MLSDS products composed of limited kinds of amino acid. It has been regarded that such a peptide can be synthesized only by 3NP method. A mutant c-Fos library that contained only four kinds of amino acid was synthesized, which was equivalent to a library synthesized by 3NP method [26]. It was a high quality library ($F = 1.00$) (Table 2). So far, fifteen libraries with various amino acid compositions were successfully synthesized.

In order to make long ORFs, we assembled 8 units of the oligomers. Stem sequences of them did not have any stop codons. A DNA library encoding 120 amino acids plus nine 5'- and 3'-flanking semi-random di-peptides (thus, total 138 amino acids) was constructed (Fig. 3).

The diversity of the synthesized library is about 10^{16} judging from the mass (data of A_{260}) and purity (data of PAGE) of synthesized DNA. With an *in vivo* selection, there is a diversity limit by the transformation step. But with an *in vitro* selection, there is no such limitation. Thus exploration of huge sequence space by *in vitro* virus [3-5] or related techniques [28,29] will become possible, depending on the experimental cost.

Comparison of MLSDS with other methods

So far, a really random library has been generated by four methods. Other methods do not give a really random library, because they can not provide a library in which all the 20 amino acids are encoded at all sites. A comparison of library quality for three methods is shown in Table 3.

An application of 3NPs method to mutagenesis of antibodies [27] or coiled-coils [30] gave good results. Twenty kinds of 3NPs mean one codon per one amino acid, but the codons are degenerate. Thus 3NPs method makes many tRNAs useless. The translation efficiency was calculated based on the codon usage, giving maximum 4-fold decrease in *Triticum aestivum*. It was reported that the reaction efficiency of 3NPs was far from uniform. The sequence data of synthesized DNA using an equimolar mixture of 19 kinds of 3NPs (without Cys) showed 12-fold (maximum) difference in composition [27] or more [12]. The coupling yield was affected by the mixing ratio

Table 3: The comparison between the target amino acids composition and the actual composition of various libraries.

A. Uniform Type			
Method	3NPs	mRNA display	MLSDS
Reference, Library	[29], 96T λ 6	[18], Random	This work, No Cys
Target value	Equimolar 19 amino acids	Equimolar 20 amino acids	Equimolar 19 amino acids
Correlation coefficient [a]	0.43	0.56	0.66
Sum of absolute errors [b] [%]	54.5	23.7	28.1
Percentage of stop codon	ND [c]	0.46	0.00
Percentage of cassettes containing stop codon	ND [c]	8.33	0.00
B. Doping Type			
Method	3NPs	mRNA display	MLSDS
Reference, Library	[29], 100z	[18], β-Cassette polar	This work, c-Fos, e', g'-
Target value	Biased	Polar amino acid	Biased
Correlation coefficient [a]	0.94	0.56	1.00
Sum of absolute errors [b] [%]	42.4	73.7	5.66
Percentage of stop codon	ND [c]	0.39	0.00
Percentage of cassettes containing stop codon	ND [c]	6.25	0.00

The values were calculated using the data of full-length (without deletion) libraries. Values of 3NPs and mRNA display method were calculated with some assumptions. In mRNA display method, the value of stop codons were estimated from data described in the article and some assumptions; one cassette did not have two or more stop codons and stop codons appeared equally in every position. The target composition of β-Cassette polar is assumed that polar amino acids are equi-molar and the others are 0%.

[a]Correlation coefficient is calculated using target and actual amio acids composition including stop codons contribution. And it is assumed that stop codons were not appeared in full-length libraries of 3NPs method.

[b] Sum of absolute values of difference between each target and actual amino acid.

[c] ND is abbreviation of no data.

of 3NPs and by the context, showing 8-fold (maximum) difference for the same 3NP [27]. Thus it will be difficult to correct reaction efficiencies by adjusting the mixing ratio. The correlation coefficient between the target composition and the actual composition was about 0.4 (for uniform 19 kinds of amino acids) [27] (Table 3). Dimer-phosphoramidites [17] method is a variation of 3NPs method, using pre-synthesized amidites, and had the same problems. In fact, the bias was observed [17].

A pre-selecting method using an mRNA display [15] was fruitful in evolutionary protein design. Novel peptide aptamers were evolved starting from a long ORF random library [31,32]. But this method could not remove all the stop codons. It gave limited library diversity. This method has low flexibility in amino acid composition. For example it is difficult to generate a "Uniform except [Met] = 0" library. The correlation coefficient between the target composition and the actual composition were not so high (Table 3).

The Y-Ligation Block Shuffling (YLBS) method [9] has high potentiality in the evolutionary design of peptides. It has problems on deletion and reaction bias of RNA ligase. MLSDS produced libraries with high quality as shown in Table 3. Above-mentioned problems are not so severe for MLSDS method, because it uses only standard phosphoramidites and is free from any biochemical bias such as in mRNA display and in YLBS. It was reported that the difference in the reaction efficiency of equimolar mixture of four kinds of mono-phosphoramidites was only about 1–5 % [33,34]. MLSDS can create any specific amino acid composition as same as 3NP method, and a MLSDS library is made at lower cost than that made with other methods.

Conclusions

We applied the split synthesis to oligodeoxyribonucleotide synthesis and developed a new procedure, Multi-Line Split DNA Synthesis (MLSDS), based on the synthesis of designed codon mixtures using three-line DNA synthesizers. MLSDS can make various "rational" libraries of huge diversity with few stop codons by using bioinformatics databases. Combination of an MLSDS library with a screening method for huge diversity will accelerate the protein evolution *in vitro*.

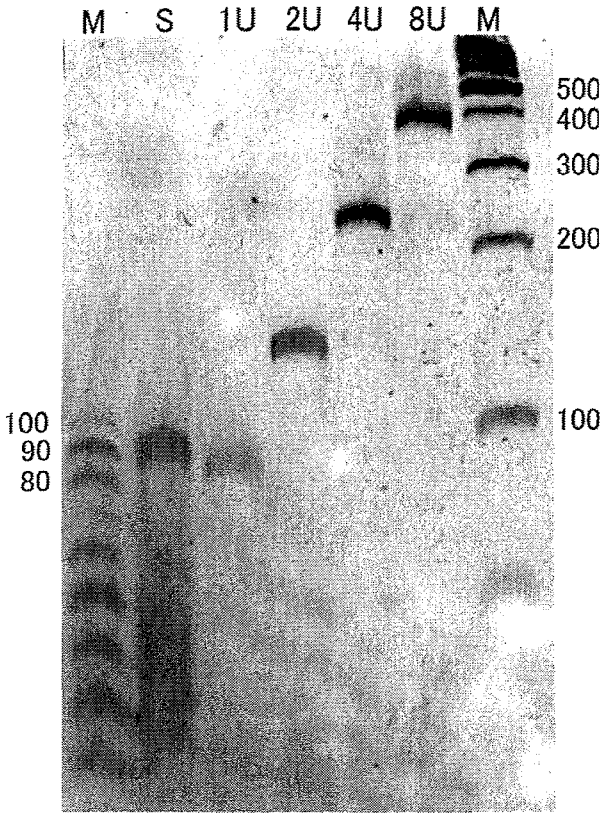


Figure 3
Production of 2×, 4×, and 8× length library. Lane-M: Size marker. 10 base ladder (left) and 100 base ladder (right). Lane-S: Initial synthesized library (87 mer). Lane-1U: Initial single-unit library (94 mer) made by PCR amplification. Lane-2U: Double-unit library (138 mer) made by ligation product of digested 1U. Lane-4U: quadruple-unit library (240 mer) Lane-8U: octuple-unit library (444 mer). PAGE was done with 5 % polyacrylamide, TAE buffer (pH 8.0), 8 M urea, 65°C and stained with SybrGreen I and visualized with a fluorescence imager (Bio-Rad FX).

Methods

A random MLSDS library was synthesized as follows. A standard DNA synthesis method was used in three lines of DNA synthesizer running in parallel. The randomized regions were combinatorialy synthesized in codon units. Triplet codons were synthesized separately in the three synthesizers as an elongation reaction of oligonucleotides on beads made of controlled pore glass (CPG). CPG beads were mixed together manually, and then splitted again into three reaction tubes manually and the next triplet codons were synthesized (Fig. 1).

The sequence of a 87 mer library was 5'-GAT GAG GCG AAG ACG *NAC TGS* $(123/456/789)_{15}$ *NAC TGS* GAG GCT GGC TGC CAC-3', where N and S denote A/T/G/C and G/C, respectively. The A:T:G:C mixing ratio in each letter of three codon groups 123, 456, and 789 was shown in Table 1. These values were calculated as described below. Both flanking regions contain the recognition sequences of type-IIs restriction enzymes *Bbs*I and *Bbv*I, respectively. In order to make longer sequences, we ligated 2 to 8 units of oligomers at the cohesive ends (the underlined sequences shown above) generated by the restriction enzyme treatment. The assembly method was as described in Ref. [16]. The italicized sequence shown above represents the assembly unit (random region of 45 bp and flanking semi-random linking region of (6+6)/2 bp).

The synthesized DNA libraries were amplified by PCR using KOD Dash polymerase (TOYOBO), inserted into pCR2.1TOPO vector (Invitrogen) and cloned, avoiding cloning bias. The clones were sequenced.

Computer calculations to determine the optimum molar mixing ratio of four bases in the codon synthesis step were performed by using Mathematica (Wolfram Research). We made a GA program for this purpose. Firstly, the target amino acid composition, $\mathbf{p}_T = (p_{T1}, p_{T2}, p_{T3},..., p_{T21})$, was established, where normally $p_{T21} = 0$ for stop codons. Secondly, we calculated an expected amino acid (plus stop codons) composition, $\mathbf{p} = (p_1, p_2, p_3,.., p_{21})$, from the molar mixing ratio of the bases, $x = (x_1, x_2, x_3,....., x_{12L})$, where $12L$ is equal to 4(number of bases) × 3(number of codon letters) × L(number of synthesizer-line). For example, the mixtures for the first letter and the second letter of the first DNA synthesizer have the molar mixing ratio [A]: [T]: [G]: [C] = $x_1 : x_2 : x_3 : x_4$ and $x_5 : x_6 : x_7 : x_8$, respectively. And for example, when $L = 3$, the expected alanine composition p_1 is given by:

$$p_1 = \frac{1}{3}\left(\frac{x_3 x_8}{(x_1+x_2+x_3+x_4)(x_5+x_6+x_7+x_8)} + \frac{x_{15}x_{20}}{(x_{13}+x_{14}+x_{15}+x_{16})(x_{17}+x_{18}+x_{19}+x_{20})} + \frac{x_{27}x_{32}}{(x_{25}+x_{26}+x_{27}+x_{28})(x_{29}+x_{30}+x_{31}+x_{32})}\right) \quad (1)$$

for the full-length sequence without deletion.

We solved an integer-programming problem (6-valued $12L$-dimensional optimization problem) having the solution x_i as integer (0,1,2,3,4,5). The reason for 6-digits "integer" was to simplify the DNA synthesizer handling and also to simplify the calculation. As the fitness F of x in the GA, we took a correlation coefficient between the expected (or designed) amino acid composition and the target amino acid composition:

$$F\{x\} = \{N\mathbf{p}_T \cdot \mathbf{p} - \sum_{i=1}^{N} p_{T,i} \sum_{i=1}^{N} p_i\} / \left\{(N\mathbf{p}_T{}^2 - (\sum_{i=1}^{N} p_{T,i})^2)(N\mathbf{p}^2 - (\sum_{i=1}^{N} p_i)^2)\right\}^{1/2}, \quad (2)$$

where $N = 21$ for our normal case. The optimum x, which gave the maximum fitness F, was calculated using a simple GA program.

It was reported the deletion rate during a general oligonucleotide synthesis is about 0.5% per coupling [20], and our data (about 0.3% per coupling) were compatible with this value. We incorporated the effect of the single nucleotide deletion into the GA calculation. We considered only the affect of a point deletion in a synthesized oligonucleotide because the deletion rate is low enough. When a point deletion occurs in the 5' constant region, all the amino acids in the random region are the frame shifted ones. When the event occurs at the i-th site of the random region, it affects the composition in the all downstream from the i-th site, and so on. We incorporated all these effects into the calculation of the composition. Details are described in Additional file 1.

Authors' contributions

IT conceived of this specific study and participated in its design and coordination. SS carried out GA calculation. SU carried out the sequence analysis and the making of the assembled longer sequence. YH conceived of general background and mathematical detail.

Abbreviations

MLSDS, multi-line split DNA synthesis; ORF, open reading frame; CPG, controlled pore glass; PAGE, polyacrylamide gel electrophoresis; GA, genetic algorithm; 3NPs, trinucleotide phosphoramidites.

Additional material

Additional File 1

In the additional WORD file (MLSDS22AdditionalFile.doc), the detail of calculation method is described for the the expected amino acids composition $p = (p_{Ala}, p_{Arg}, \ldots\ldots, p_{term})$ *considering the single nucleotide deletion.*

Click here for file

[http://www.biomedcentral.com/content/supplementary/1472-6750-4-19-S1.doc]

Acknowledgements

We thank Dr. T. Aita and Dr. T. Sasaki for useful advise and discussions, and Dr. J. Hattori for critical reading of the manuscript.

References

1. Lam KS, Salmon SE, Hersh EM, Hruby VJ, Kazmierski WM, Knapp RJ: **A new type of synthetic peptide library for identifying ligand-binding activity.** *Nature* 1991, **354**:82-84.
2. Smith GP: **Filamentous fusion phage: novel expression vectors that display cloned antigens on the virion surface.** *Science* 1985, **228**:1315-1317.
3. Roberts RW, Szostak JW: **RNA-peptide fusions for the in vitro selection of peptides and proteins.** *Proc Natl Acad Sci USA* 1997, **94**:12297-12230.
4. Tabuchi I, Soramoto S, Nemoto N, Husimi Y: **An in vitro DNA virus for in vitro protein evolution.** *FEBS Lett* 2001, **508**:309-312.
5. Tabuchi I: **Next-generation protein-handling method: puromycin analogue technology.** *Biochem Biophys Res Commun* 2003, **305**:1-5.
6. Arkin AP, Youvan DC: **Optimizing nucleotide mixtures to encode specific subsets of amino acids for semi-random mutagenesis.** *Biotechnology* 1992, **10**:297-300.
7. LaBean TH, Kauffman SA: **Design of synthetic gene libraries encoding random sequence proteins with desired ensemble characteristics.** *Protein Sci* 1993, **2**:1249-1254.
8. Tomandl D, Schober A, Schwienhorst A: **Optimizing doped libraries by using genetic algorithms.** *J Comput Aided Mol Des* 1997, **11**:29-38.
9. Kitamura K, Kinoshita Y, Narasaki S, Nemoto N, Husimi Y, Nishigaki K: **Construction of block-shuffled libraries of DNA for evolutionary protein engineering: Y-ligation-based block shuffling.** *Protein Eng* 2002, **15**:843-853.
10. Shiba K, Takahashi Y, Noda T: **Creation of libraries with long ORFs by polymerization of a microgene.** *Proc Natl Acad Sci U S A* 1997, **94**:3805-3810.
11. Sondek J, Shortle D: **A general strategy for random insertion and substitution mutagenesis: substoichiometric coupling of trinucleotide phosphor-amidites.** *Proc Natl Acad Sci U S A* 1992, **89**:3581-3585.
12. Virnekas B, Ge L, Pluckthun A, Schneider KC, Wellnhofer G, Moroney SE: **Trinucleotide phosphoramidites: Ideal reagents for the synthesis of mixed oligonucleotides for random mutagenesis.** *Nucleic Acids Res* 1994, **22**:5600-5607.
13. Kayushin AL, Korosteleva MD, Miroshnikov AI, Kosch W, Zubov D, Piel N: **A convenient approach to the synthesis of trinucleotide phosphoramidites – synthons for the generation of oligonucleotide/peptide libraries.** *Nucleic Acids Res* 1996, **24**:3748-3755.
14. Ono A, Matsuda , Zhao J, Santi DV: **The synthesis of blocked triplet-phosphoramidites and their use in mutagenesis.** *Nucleic Acids Res* 1995, **23**:4677-4682.
15. Cho G, Keefe AD, Liu R, Wilson DS, Szostak JW: **Constructing high complexity synthetic libraries of long ORFs using in vitro selection.** *J Mol Biol* 2000, **297**:309-319.
16. Glaser SM, Yelton DE, Huse WD: **Antibody engineering by codon-based mutagenesis in a filamentous phage vector system.** *J Immunol* 1992, **149**:3903-3913.
17. Neuner P, Cortese R, Monaci P: **Codon-based mutagenesis using dimer-phosphoramidites.** *Nucleic Acids Res* 1998, **26**:1223-1227.
18. Cormack BP, Valdivia RH, Falkow S: **FACS-optimized mutants of the green fluorescent protein (GFP).** *Gene* 1996, **173**:33-38.
19. Gaytan P, Osuna J, Soberon X: **Novel ceftazidime-resistance beta-lactamases generated by a codon-based mutagenesis method and selection.** *Nucleic Acids Res* 2002, **30**:e84.
20. Hecker KH, Rill RL: **Error analysis of chemically synthesized polynucleotides.** *Biotechniques* 1998, **24**:256-260.
21. Proba K, Worn A, Honegger A, Pluckthun A: **Antibody scFv fragments without disulfide bonds made by molecular evolution.** *J Mol Biol* 1998, **275**:245-253.
22. Aita T, Iwakura M, Husimi Y: **A cross-section of the fitness landscape of dihydrofolate reductase.** *Protein Eng* 2001, **14**:633-638.
23. Hirao I, Ohtsuki T, Fujiwara T, Mitsui T, Yokogawa T, Okuni T, Nakayama H, Takio K, Yabuki T, Kigawa T, Kodama K, Yokogawa T, Nishikawa K, Yokoyama S.: **An unnatural base pair for incorporating amino acid analogs into proteins.** *Nat Biotechnol* 2002, **20**:177-181.
24. Klapper MH: **The independent distribution of amino acid near neighbor pairs into polypeptides.** *Biochem Biophys Res Commun* 1977, **78**:1018-1024.
25. King JL, Jukes TH: **Non-Darwinian evolution.** *Science* 1969, **164**:788-798.
26. Pelletier JN, Arndt KM, Pluckthun A, Michnick SW: **An in vivo library-versus-library selection of optimized protein-protein interactions.** *Nat Biotechnol* 1999, **17**:683-690.
27. Knappik A, Ge L, Honegger A, Pack P, Fischer M, Wellnhofer G, Hoess A, Wolle J, Pluckthun A, Virnekas B: **Fully synthetic human combinatorial antibody libraries (HuCAL) based on modular**

consensus frameworks and CDRs randomized with trinucleotides. *J Mol Biol* 2000, **296**:57-86.
28. Tabuchi I, Soramoto S, Suzuki M, Nishigaki K, Nemoto N, Husimi Y: **An efficient ligation method in the making of an *in vitro* virus for *in vitro* protein evolution.** *Biol Proced Online* 2002, **4**:49-54.
29. Husimi Y, Aita T, Tabuchi I: **Correlated Flexible Molecular Coding and Molecular Evolvability.** *J Biol Phys* 2002, **28**:499-507.
30. Arndt KM, Pelletier JN, Muller KM, Alber T, Michnick SW, Pluckthun A: **A heterodimeric coiled-coil peptide pair selected in vivo from a designed library-versus-library ensemble.** *J Mol Biol* 2000, **295**:627-639.
31. Wilson DS, Keefe AD, Szostak JW: **The use of mRNA display to select high-affinity protein-binding peptides.** *Proc Natl Acad Sci USA* 2001, **98**:3750-3755.
32. Keefe AD, Szostak JW: **Functional proteins from a random-sequence library.** *Nature* 2001, **410**:715-718.
33. Jellis CL, Cradick TJ, Rennert P, Salinas P, Boyd J, Amirault T, Gray GS: **Defining critical residues in the epitope for a HIV-neutralizing monoclonal antibody using phage display and peptide array technologies.** *Gene* 1993, **137**:63-68.
34. Hermes JD, Parekh SM, Blacklow SC, Koster H, Knowles JR: **A reliable method for random mutagenesis: the generation of mutant libraries using spiked oligodeoxyribonucleotide primers.** *Gene* 1989, **84**:143-151.