

ORIGIN OF GENOMIC DNA: DISCUSSION FROM REVERSE-TRANSCRIPTION AND EXPANSION OF
REPETITIVE OLIGONUCLEOTIDES

Takanori Miura

Taiko Pharmaceutical Co., Ltd., Uchihonmachi 3-34-14 Suita Osaka 564-0032 JAPAN

TEL +81-6-6382-1023, FAX +81-6-6382-3084

E-mail: takanori@myrealbox.com

(Received 25 October, 2002 Accepted 3 February 2003)

(Abstract)

Genetic information of organisms is preserved within sequences of nucleotides in genomic DNAs, which are possible to be replicated by DNA polymerase semiconservatively. The sequences of the DNA molecules in organisms are variant, but long. Even the DNA of *Mycoplasma genitalium*, which has a smallest DNA among all types of organisms, is of 580-kilobase pairs in length. However, the long DNA can be synthesized with neither some prebiotic experiments nor the present DNA synthesizing techniques. Only oligoribonucleotides 55mer are possible to be synthesized in the presence of a montmorillonite clay under prebiotic conditions. How is the long DNA synthesized in the primordial environment? Ohno proposed the model that genes of the present organisms emerged from short repetitive oligoribonucleotides. Several types of DNA polymerase are capable to synthesize long repetitive DNAs with more than 50-kilobase pairs in length from short oligonucleotides such as $(TA)_n$, $(TG)_n$, $(CAG)_n$, $(TAGG)_n$, $(TTAGGG)_n$, $(TACATGTA)_n$, and $(AGATATCT)_n$ by conventional enzymatic reactions. Furthermore, telomerase, which catalyzes the synthesis of the telomere of eucaryotic genome ends, can synthesize the complementary repetitive DNA $(TTGGGG)_n$ in the presence of a RNA template by the reverse-transcriptional activity. There are various repetitive sequences such as $(GT)_n$, $(CAG)_n$, $(GGAAT)_n$, and $(TTAGGC)_n$ in the genomic DNA of different organisms containing Bacteria and Eucarya. Based on these examples, it is deduced that short repetitive DNA should have been synthesized from short RNA in the presence of oligopeptides catalysts or ribozymes, which have primordial reverse transcriptase-like activity; genomic DNA should have emerged from long repetitive DNAs elongated from the short repetitive DNA in the presence of primordial DNA polymerase-like oligopeptides or ribozymes. I discuss the plausibility of these ideas, proposing a model for the origin of genomic DNA from the reverse-transcription and expansion of repetitive oligonucleotides.

(Keyword)

repetitive oligonucleotide, tandem repeat, DNA polymerase, telomerase, genomic DNA, DNA elongation
chemical evolution of DNA

ゲノム DNA の起源
：短鎖反復核酸の逆転写および伸長合成からの考察
三浦 孝典

大幸薬品株式会社

〒564-0032 吹田市内本町 3-34-14

TEL 06-6382-1023, FAX 06-6382-3084

E-mail: takanori@myrealbox.com

(要旨)

現存生物の遺伝情報のほとんどはゲノム DNA に記録されており、これは DNA ポリメラーゼにより半保存的に複製されている。その長さは、最小ゲノム DNA を有するマイコプラズマ・ゲニタリウムでさえ、58 万塩基対と非常に長い。このような長鎖 DNA は、現在の DNA の化学合成技術をもってしても合成できないが、ましてや生命誕生以前の環境を模倣した化学進化モデル反応では実現されていない。唯一、55 塩基長のオリゴリボヌクレオチドが粘土鉱物存在下で無生物的に生成することが見いだされたのみである。はたして、ゲノム DNA はどのようにして無生物的に非常に長い DNA となったのであろうか？ 大野は、現存生物の遺伝子は短鎖反復核酸が伸長した反復 RNA から誕生したモデルを提唱した。DNA ポリメラーゼは、短鎖反復 DNA を、たとえば(TA)_n、(TG)_n、(CAG)_n、(TAGG)_n、(TTAGGG)_n、(TACATGTA)_n、(AGATATCT)_nなどの長鎖反復 DNA へ伸長することが知られている。また、テロメラーゼは鋳型 RNA に相補的な反復 DNA(TTAGGG)_n を逆転写的に合成することが報告されている。そして、細菌や真核生物を含む現存生物のゲノム DNA 中には、たとえば(GT)_n、(CAG)_n、(GGAAT)_n、(TTAGGC)_nなどの反復 DNA が多数存在している。これらの知見から筆者は、まず、オリゴリボヌクレオチドとオリゴペプチドが無生物的に合成され、それらの分子の中で逆転写活性を持つオリゴペプチドかまたはオリゴヌクレオチド(リボザイム)が反復オリゴリボヌクレオチドを鋳型として反復オリゴデオキシリボヌクレオチドを逆転写し、反復 DNA を合成したのではないかと仮説をたてた。そして、この反復 DNA が、DNA ポリメラーゼ活性を持つ短鎖ペプチドまたはリボザイムにより、長鎖反復 DNA に伸長し、ゲノム DNA の起源になったのではないかと考えた。筆者は、これらの考えを検証した上で、反復核酸の逆転写と伸長を元にしたゲノム DNA の起源のモデルを提唱する。

1. はじめに

地球の誕生は約 46 億年前と推定されており、生物の誕生は最古の微生物化石の発見から 38 億年前と推定されている[1, 2]。誕生後間もない原始生物は、現在の生物と同様にデオキシリボ核酸(DNA)、リボ核酸(RNA)、タンパク質を有しており、その後突然変異と自然淘汰の繰り返しにより、現在の生物種へと進化してきたであろうという考えが主流であった [1]。しかし、RNA が酵素活性(なお、酵素活性をもつ RNA をリボザイムと呼ぶ)と遺伝情報の担い手の両方を兼ね備えていることが発見され、RNA 単独で生命活動が

行われていたという RNA ワールド仮説が最近広まっている[3]。いずれにせよ、現存生物は遺伝情報としてゲノム DNA を有しており、ある段階で合成されたことは疑いない。多くの生物のゲノム DNA は、それぞれのゲノムプロジェクトにより急速に塩基配列解読が進んでおり、その正確な塩基長も解明されてきている。Table 1 にいくつかの例を示す。細菌、動物、および植物の一例である大腸菌(*Escherichia coli*)[4]、ヒト(*Homo sapiens*)[5]、およびシロイソナズナ(*Arabidopsis thaliana*)[6]のゲノム DNA 長は、それぞれ約 400 万塩基対、約 31 億塩基対、および約 1 億塩基対であり、また、現存する生物の中で最小ゲノムを有するマイコ

Table 1. Typical genome sizes and gene amounts of some organisms

Species	genome size	gene number
<u>Bacteria</u>		
<i>Mycoplasma genitalium</i>	580 kb	484
<i>Buchnera</i> sp.	640 kb	564
<i>Escherichia coli</i> K12	4.64 Mb	4289
<i>Bacillus subtilis</i>	4.21 Mb	4100
<i>Pseudomonas aeruginosa</i>	6.3 Mb	5565
<i>Synechocystis</i> sp. PCC6803	3.57 Mb	3169
<u>Archaea</u>		
<i>Methanococcus jannaschii</i>	1.66 Mb	1715
<i>Pyrococcus harikoshii</i>	1.74 Mb	1979
<i>Aeropyrum pernix</i>	1.67 Mb	2694
<i>Halobacterium</i> sp.	2 Mb	2058
<u>Eucarya</u>		
<i>Saccharomyces cerevisiae</i>	12 Mb	6286
<i>Caenorhabditis elegans</i>	97 Mb	18000
<i>Drosophila melanogaster</i>	120 Mb	13000
<i>Arabidopsis thaliana</i>	125 Mb	25498
<i>Homo sapiens</i>	3 Gb	30000

(kb: 1000 bases; Mb: 1000000 bases; Gb: 1000000000 bases)

プラズマ・ゲニタリウム(*Mycoplasma genitalium*)でさえ、58万塩基対と非常に長いゲノムDNAを持っている[7]。なお、生命誕生に必要な最小のゲノムDNAの長さは必須遺伝子の数(517個)からおよそ40万塩基対である、と推定した報告もある[8]。そこで、このような長大なゲノムDNAは、どのような機構で合成されたのか、という疑問を感じるのは著者だけではないであろう。

リボヌクレオチド類似体を出発材料として短鎖リボヌクレオチド(RNA)が無生物的に合成されることは、Orgel[9]、澤井[10]、Ferris[11]らにより多数報告された。合成された短鎖RNAの長さは、Ferrisらの粘土存在下での無生物的合成反応の55塩基長が最長であり[12]、長大な核酸の無生物的な合成はいまだ観察されていない。長大な核酸合成のモデルとしては、単純な短鎖反復RNAの無生物的な伸長合成による長鎖RNA生成が大野乾により提唱されており[13]、長鎖DNAの前生物的合成モデルは提示されていなかった。そこで筆者は、ゲノムDNA誕生における長鎖DNAの前生物的合成モデルを考えた。

本稿は、まず第2章でリボヌクレオチド類似体から短鎖RNAの無生物的合成についての過去の知見を分析し、ゲノムDNA誕生における問題点を述べる。続く第3章では、大野が提唱した単純な短鎖反復RNAの無生物的な伸長合成による長鎖反復RNA遺伝子の

誕生仮説[13]を分析し、ゲノムの誕生における補足事項を考察する。第4章では、現存生物の酵素が有しているRNAからDNAへの逆転写反応、短鎖DNAから長鎖DNAの伸長合成反応を分析し、ゲノムDNA誕生におけるこれらの反応の重要性を考える。そして第5章では、第2章から第4章の知見から導き出される事柄を記載し、短鎖RNAから長鎖反復DNAの無生物的合成反応による初期ゲノムDNA誕生モデルを提案する。

2. 核酸およびタンパク質の前生物的合成モデルの概観とその問題点の分析

短鎖核酸の前生物的合成モデルとして1960年代後半から現在に至るまで積極的に行われている。本章では、モノヌクレオチドおよび短鎖核酸の無生物的合成実験を分析し、その問題点を述べる。

DNAやRNAのような核酸は、糖、塩基、リン酸から構成されている。糖の一種、リボースは、ホルムアルデヒドから無生物的に合成され[14]、塩基の一種、アデニンは、シアン酸水素とアンモニアから無生物的に合成された[15]。糖とアデニンの重合体、アデノシンとデオキシアデノシンに関しても無生物的に重合できることが示された[16]。ヌクレオシドのリン酸化は、尿素と無機リン酸の存在下でおこり[17]、さらに反応

を進めると、ヌクレオシド 環状 2', 3'-リン酸(nucleoside cyclic 2', 3'-phosphate, cpNp)やヌクレオシド 2 リン酸、ヌクレオシド 5'-ポリリン酸(nucleoside 5'-polyphosphate, polypN)が合成された[18]. この polypN とアミンを乾燥固化させると ImpN が無生物的に合成され [19], これは次段落以降に示す核酸モノマーからの無生物的重合反応に多用された. 以上の研究結果から, モノリボヌクレオチドは前生物的に存在していたことが推定された.

続いてモノリボヌクレオチドの重合による短鎖 RNA の無生物的合成について述べる. cpNp の一種, アデノシン 環状 2', 3'-リン酸(adenosine cyclic 2', 3'-phosphate, cpAp)は, 固相反応条件下で反応させると 2 塩基長以上の短鎖 RNA を生じた[20]. 水溶液の反応条件では, ImpN の一種, アデノシン 5'-リン酸イミダゾリド(adenosine 5'-phosphorimidazolide, ImpA)が 2 価金属イオンの存在する中で 5 塩基長まで無生物的に重合することや [10,21], 同様の重合反応でシチジン 5'-リン酸イミダゾリド(cytidine 5'-phosphorimidazolide, ImpC)が 6 塩基長まで合成できることが示された[22]. これらの結果から, 核酸モノマーを出発材料として 6 塩基長までではあるが短鎖 RNA が前生物的に合成されていたことが推定された.

次に鋳型 RNA を加えた反応条件でのモノリボヌクレオチドから短鎖 RNA の無生物的合成についてまとめる. 鋳型 RNA としてポリヌクレオチド(C)_nを加えたとき, グアノシン 5'-リン酸イミダゾリド (guanosine 5'-phosphorimidazolide, ImpG) が 30 塩基長の(G)₃₀まで無生物的に重合でき, 飛躍的に短鎖核酸の鎖長が増加した [9]. 活性化ヌクレオチドを ImpN からヌクレオシド 5'-リン酸-2-メチルイミダゾリド(nucleoside 5'-phosphor-2-methylimidazolide)に変更したとき, 合成量の増加, 3'-5'リン酸ジエステル結合の短鎖 RNA が合成される割合の増加, 短鎖 RNA の塩基長の増加がみられた. 例えば, ポリヌクレオチド(C)_nの鋳型に対して, グアノシン 5'-リン酸-2-メチルイミダゾリド (guanosine 5'-phosphor-2-methylimidazolide, MeImpG)を用い, 縮合剤は用いずに, Mg²⁺の存在下で反応させたとき, 3'-5'リン酸ジエステル結合をもつ最長 50 塩基長の短鎖 RNA が合成された[23]. その他には, 鋳型 RNA (C, U)_nに対する A と G を含有する 13 塩基長のオリゴヌクレオチドの複製[24], ランダムポリヌクレオチド

(C,G)_nによる 12 塩基の G と C 含有オリゴヌクレオチドの複製[25]などの報告もある. これらの実験より, 50 塩基長までの短鎖 RNA が無生物的に合成可能なが示されたが, 鋳型 RNA がどのようにして生成されたかという疑問が依然残る.

Ferris のグループは, 鋳型 RNA の代わりに粘土鉱物の一種モンモリロナイトを添加して, ImpA から無生物的に短鎖 RNA が合成できることを示した[11]. さらに研究を重ね, 最長 55 塩基長の 3'-5'リン酸ジエステル結合を持つ短鎖 RNA が無生物的に合成できた [12]. 反応論的解析より, このモンモリロナイトを添加した短鎖 RNA の無生物的合成系は, (C)_nの鋳型に対する MeImpG 重合による短鎖 RNA の無生物的合成系よりも反応速度定数が大きかった[26]. 鉱物は ImpA 吸着と反応の場を提供しており, 太古の地球上にも存在したと思われる. したがって, 前生物的に 55 塩基長までの短鎖 RNA の合成が原始地球条件下で可能なことが支持される.

次にモノデオキシリボヌクレオチドの重合による短鎖 DNA の無生物的合成に関して考察する. DNA の糖を構成するデオキシリボースは, 無生物的には合成しにくいことが報告された[27]. また, デオキシリボヌクレオチドの類似体デオキシアデノシン 5'-リン酸イミダゾリド (deoxyadenosine 5'-phosphorimidazolide, ImpdA)は無生物的な重合反応がほとんどおこらないことも示された[22]. 一方で, 高温乾燥という特殊な条件下でデオキシチミジン 5'-3'リン酸(dTTP)が縮合剤 [cyanamide か 4-amino-5-imidazole carboxamide (AICA)]を用いて 4 から 8 塩基長まで重合でき [28,29], 潮だまりを連想させる固化溶解繰り返しシステムで, 縮合剤存在下で 4 塩基長の短鎖 DNA が合成できることも報告された[30]. しかし, 縮合剤が特殊な点など前生物的合成モデルとしては受け入れがたく, なおこれら一連の研究は以後報告がない. したがって, 短鎖 DNA の無生物的合成は特殊なケースでは可能であるが, 前生物的合成モデルとしての現実性は乏しいことが示唆された.

以上まとめると, 短鎖核酸の無生物的合成に関して, 溶液中ではない特殊な条件下で短鎖 DNA の合成もおこるが, しかし, 前生物的核酸合成モデルとして有望な活性化デオキシリボヌクレオチドをもちいた条件下では短鎖 DNA の合成はほとんどおこらない. 短鎖 RNA

の合成は粘土鉱物を加えると比較的長い 55 塩基長の短鎖 RNA の合成が可能であった。鋳型 RNA の存在下においても、縮合剤を使わない極めて単純な反応系で 50 塩基長の短鎖 RNA が合成されており、短鎖 RNA の無生物的合成は容易である。したがって、短鎖核酸の前生物的合成モデルとしては、短鎖 RNA が最初に合成されたと考えられる。

3. 大野乾が提唱した短鎖反復核酸の伸長による遺伝子誕生モデルとそのモデルの分析

1980 年代に入り、遺伝子の塩基配列が蓄積される中、大野乾は、遺伝子およびその翻訳後のタンパク質の中に繰り返し出現する(周期性のある)塩基配列とそれを翻訳したアミノ酸配列が多く存在することを見つけた。そこで、異なる生物種の血清アルブミン属のタンパク質、ウシ血清アルブミンとマウス α -フェトプロテインの塩基配列とアミノ酸配列を比較し、繰り返し出現する配列が 2 種類あることを突き止め、これらを原始配列であると推論した [31]。その後、主要組織適合抗原遺伝子 [32]、ヒトのガン遺伝子 *c-myc* とニワトリの *v-myc* 遺伝子 [33]、マウス免疫グロブリン *Ig C_{μH}* 遺伝子 [34]、 β_2 -ミクログロブリン遺伝子 [35]、ニワトリのガン遺伝子 *c-src* とウイルスの相同遺伝子 *v-src* [36]、マラリア原虫のスポロゾイド抗原遺伝子 [37]、ブタのムスカリン・アセチルコリンレセプター遺伝子 [38]、マスの H1 ヒストン遺伝子 [13] でそれぞれ個別に原始配列を推定し、報告している。例えば、スポロゾイド抗原の場合、36 塩基長 GGACAACCACAAGCACAAGGAGA-TGGAGCAAATGCA が 12 回反復しており、原始配列と推定された。しかし、この配列の前半は、さらに 6 塩基長の類似配列 ACAACC, ACAAGC, ACAAGG を含み、短い原始配列に分割できることがわかる。すなわち、長い反復配列の内部に短い反復配列を含む構造であった。このような観察結果から、大野は反復配列に着目し、遺伝子の誕生および進化に関して優れた考察を残した [13,37]。

大野は、遺伝子が誕生するために必要な 3 つの要件として、(1)長い読みとり枠(open reading frame)を容易に形成できる、(2) α ヘリックスや β シートなどの 2 次構造を容易に形成できる、(3)塩基配列の欠失、挿入、

置換などの突然変異に対して強い読みとり枠を有している、ことを考えた。まず、彼は (1) の要件に関して 300 塩基長の RNA の翻訳できる確率、つまり 100 アミノ酸残基のタンパク質遺伝子を生成できる確率をランダム配列からの場合と反復配列からの場合で算出し、比較を行った。現存生物の遺伝暗号は、1 開始コドン、3 終止コドン、61 アミノ酸コドンである。そこで、ランダム配列から遺伝子を生成できる確率は、 $(1/64) \times (61/64)^{99} = 1.35 \times 10^{-4}$ 、9 塩基長の反復配列からの確率は、 $(3/64) \times (61/64)^2 = 4.23 \times 10^{-2}$ と算出した [39]。この計算結果より彼は、反復配列が (1) の要件においてランダム配列より勝っていると考えた。しかし、長い読みとり枠が形成されただけで、機能のあるタンパク質を作る遺伝子となるわけではない。 α ヘリックスや β シートなどの 2 次構造を形成できるアミノ酸配列になりうる塩基配列が必要となる。例えば、アルギニン(Arg)、アスパラギン(Asn)、グルタミン(Gln)、グルタミン酸(Glu)などが規則性を持ち重合したペプチド断片は、 α ヘリックス構造を形成しやすい。そこで彼は、10 塩基長の DNA 鎖 ATCAGGAGGA を基本単位として持つ反復配列は、3 回反復させた 30 塩基長の DNA (ATCAGGAGGA)₃ から 12 残基長のペプチド鎖 Ile-Arg-Arg-Asn-Gln-Glu-Glu-Ser-Gly-Gly が翻訳でき、 α ヘリックス構造を形成しやすくなると考えた。これは、ランダム塩基配列から 2 次構造を形成するより明らかに容易であり、(2)の要件においても反復配列が優れていることを主張した。また大野は、反復 DNA 配列の反復単位が 3 の倍数でない方が長いペプチド鎖を形成できると考えた。つまり、3 塩基で 1 アミノ酸をコードするので、12 塩基長の反復単位を持つ短鎖 DNA が 3 回反復しても、4 残基長の反復単位のオリゴペプチドが 3 回反復したタンパク質しか形成できない。しかし、上の例のように 10 塩基単位長の反復配列の場合、3 回反復すると 12 残基長のタンパク質が形成でき、タンパク質残基の反復単位長の増加がみられる。この反復配列は塩基の欠失、挿入、置換による突然変異により終止コドンが発生しても、残り 2 つの読みとり枠が開かれ、長い読みとり枠が再び現れる可能性があり、読みとり枠の短小化の影響を受けにくいと大野は考えた。そこで彼は、(3)の要件においても反復配列が優れていると主張した [37]。次に大野は、遺伝子の誕生と進化に関して考察した。

核酸の前生物的合成モデルは、第 2 章で詳細に解説した。おおかた受け入れやすいモデルとしては、活性化リボヌクレオチドがモノマーからオリゴマー（短鎖リボ核酸）へ重合する Orgel と澤井によるモデル[10]、短鎖リボ核酸を鋳型として先ほどの類似体が相補的に重合する Orgel らによるモデル[9]がある。そこで彼は、5 塩基長短鎖リボ核酸 GCCAA が前生物的に合成され、連結した後 10 塩基長の反復短鎖リボ核酸 GCCAAGCCAA が合成したと仮定した。Orgel らのモデルに従えば、この短鎖リボ核酸が鋳型となり、無生物的に相補鎖 UUGGCUUGGC が合成できる。この 10 塩基対の 2 本鎖リボ核酸 GCCAAGCCAA/UUGGCUUGGC が解離して、あらためて各 3'末端側の 5 塩基ずつだけで水素結合すると、各 5'末端側が 5 塩基それぞれはみ出る。これら突出した部分を鋳型として相補鎖が複製されれば、塩基長は伸長する。この解離、再結合、複製を繰り返すことで長鎖リボ核酸が合成できると大野は考えた。この 15 塩基長 GCCAAGCCAAGCCAA は 5 残基長 Ala-Lys-Pro-Ser-Gln のオリゴペプチドに翻訳され、終止コドンを含まない。読みとり枠が 1 塩基ずつずれても、CCAAGCCAAGCCAAG から Pro-Ser-Gln-Ala-Lys, CAAGCCAAGCCAAGC から Gln-Ala-Lys-Pro-Ser が翻訳され、アミノ酸配列は同じとなり、突然変異に強い。この塩基配列は、前段落冒頭の遺伝子誕生の要件を満たす。そこで大野は、単純な直列型反復配列を有する短鎖リボ核酸が無生物的に伸長合成して長鎖リボ核酸となりこれがタンパク質の遺伝子となったのではないかと仮説を唱えた[13]。

しかし、筆者はこの仮説には実験的に示されていない多くの問題があると考え。例えば、(1)鋳型非依存的短鎖リボ核酸の連結が可能であるか、(2)数十塩基長以上のリボ核酸を無生物的に合成可能であるか、(3)原始翻訳系のモデルがない、などがそれである。なお、この説は遺伝子の誕生という観点で RNA とタンパク質を基本とした世界（RNA・タンパク質ワールド）では通用するが、現在の DNA および RNA とタンパク質の世界（DNA・RNA・タンパク質ワールド）における長鎖ゲノム DNA の誕生は示していない。そこで以後の章では、大野仮説を前提として RNA・タンパク質ワールドから DNA・RNA・タンパク質ワールドへの道筋および長鎖ゲノム DNA の誕生について考察

する。

4. 短鎖ヌクレオチドから長鎖 DNA の試験管内伸長合成

現存生物のゲノム DNA は、DNA ポリメラーゼにより半保存的に合成されている[40]。つまり、通常は 2 本鎖となっているゲノム DNA がそれぞれほどけて鋳型となり、DNA ポリメラーゼが鋳型に相補的なプライマーの 3'-OH 末端から相補的なモノヌクレオチドが重合することで、2 本鎖ゲノム DNA がコピーできる(Fig. 1a)。したがって、DNA ポリメラーゼは、通常の DNA 複製においては、あらかじめ存在する鋳型 DNA の塩基配列に納められている遺伝情報を忠実に複製するのみで、この鋳型 DNA の長さを超える DNA は合成できないことになる。しかし、反復配列という限定された配列を持つ DNA において、DNA ポリメラーゼが短鎖から長鎖へと伸長合成する試験管内実験結果が、多くのグループにより報告された(Table 2)。

1964 年の Kornberg と Khorana らは、*Escherichia coli* の DNA ポリメラーゼ [41]が、2 塩基 (TA) の反復単位を持つ 6, 8, 10, 12, 14 塩基長の (TA)₃（下付きの数字は反復単位数を表す）、(TA)₄, (TA)₅, (TA)₆, (TA)₇ 短鎖反復合成 DNA を、長鎖反復 DNA (TA)_n へ伸長合成できること報告した[42]。彼らは、この短鎖 DNA から長鎖 DNA への伸長反応モデルとして、DNA ポリメラーゼのスリップ合成を提案している (Fig. 1b)[43]。この時代には電気泳導法が確立されていなかったため、伸長した DNA の長さは確かめられていないが、粘度変化からはかなり長い DNA が生成したと推測された。その後 Khorana のグループは、(TG)₂₋₆/(AC)₂₋₆ [44]、(TAG)₄/(TAC)₃₋₄ [45]、あるいは (TATC)₃/(TAGA)₂ [46] などの 2, 3, 4 塩基長の反復単位からなる反復配列が、同 DNA ポリメラーゼによって伸長することを確かめた。これらの反応もスリップ合成と考えられている。なお、伸長した DNA の長さは、短鎖 (TA) 反復 DNA から長鎖 (TA) 反復 DNA 伸長実験と同様に確認されていない。

1992 年に Schlötterer と Tautz は、2 塩基および 3 塩基反復単位を有する多様な反復 DNA が、37°C でスリップ合成することをアガロースゲル電気泳導法で解析し、最長 700 塩基対まで伸長することを示した[47]。

Table 2. Summary of repetitive DNA elongation with some DNA polymerase

Short DNA sequence (5'→3')	Reaction temperature (°C)	DNA length elongation		DNA polymerase	Reference
		before (bp)	after (bp)		
(TA)3-7	0-45	6-14	ND	E. coli pol I	42
(TG)2-6/(CA)2-6	37	4-12	ND	E. coli pol I	44
(TTC)3-4/(AAG)3-5	37	9-15	ND	E. coli pol I	45
(TAC)3/(TAG)4-5	37	9-15	ND	E. coli pol I	45
(TAG)4/(TAC)3-4	37	9-15	ND	E. coli pol I	45
(ATG)3/(ATC)3-5	37	9-15	ND	E. coli pol I	45
(TATC)3/(TAGA)2	37	8-12	ND	E. coli pol I	46
(TAA)2/(TTAC)3-4	37	8-16	ND	E. coli pol I	46
(GT)15/(CA)9	37	18, 30	460	E. coli pol I	47
(AG)15/(TC)9	37	18, 30	700	E. coli pol I	47
(ATT)15/(AAT)9	37	27, 45	400	E. coli pol I	47
(GTT)15/(ACA)9	37	27, 45	120	E. coli pol I	47
(TCT)15/(AGA)9	37	27, 45	250	E. coli pol I	47
(TCA)15/(ATG)9	37	27, 45	110	E. coli pol I	47
(TAC)15/(TAG)9	37	27, 45	350	E. coli pol I	47
(TCC)15/(GGA)9	37	27, 45	200	E. coli pol I	47
(CCA)15/(GGT)9	37	27, 45	100	E. coli pol I	47
(TCG)15/(ACG)9	37	27, 45	50	E. coli pol I	47
(TGC)15/(GCA)9	37	27, 45	50	E. coli pol I	47
(GCC)5/(GCC)5	37	15	250	Taq pol	48
(GCC)5/(GCC)5	37	15	80	Klenow	48
(GCC)5/(GCC)5	37	15	32	Sequenase	48
(GCC)5/(GCC)5	37	15	45	HIV-RT	48
(GCC)5/(GCC)5	37	15	30	Calf pol α	48
(GCC)5/(GCC)5	37	15	40	mouse pol β	48
(AT)8	65	16	20000	Taq pol	49
(TA)9	65	18	20000	Taq pol	49
(AAT)3/(ATT)5	47	9, 15	20000	Klenow	50
(AAT)3/(ATT)5	47	9, 15	2200	Taq pol	50
(CAG)3/(CTG)5	47	9, 15	2000	Taq pol	50
(TACATGTA)6	74	48	10000	Tth pol	52
(TACATGTT)6	74	48	ND	Tth pol	52
(TACATGAA)6	74	48	ND	Tth pol	52
(TACATCTA)6	74	48	ND	Tth pol	52
(TACAAGTA)6	74	48	10000	Tth pol	52
(TACATGAT)6	74	48	ND	Tth pol	52
(TACATCAT)6	74	48	ND	Tth pol	52
(TACAACAT)6	74	48	ND	Tth pol	52
(AGGTACCT)6	74	48	10000	Tth pol	52
(AGATATCT)3	74	24	20000	Tli pol	53
(AT)24	74	48	ND	Tli pol	53
(AGGTACCT)6	74	48	ND	Tli pol	53
(TGATATCA)6	74	48	ND	Tli pol	53
(ACATATGT)6	74	48	ND	Tli pol	53
(TCATATGA)6	74	48	ND	Tli pol	53
(ACTTAAAGT)6	74	48	20000	Tli pol	53
(TATTAAGA)6	74	48	ND	Tli pol	53
(AAAGCTTT)6	74	48	20000	Tli pol	53
(TCTATAGA)6	74	48	ND	Tli pol	53
(TACATGTA)6	74	48	ND	Tli pol	53
(AACATGTT)6	74	48	ND	Tli pol	53
(TTCATGAA)6	74	48	ND	Tli pol	53
(TAGATCTA)6	74	48	ND	Tli pol	53
(TACTAGTA)6	74	48	ND	Tli pol	53
(TAAGCTTA)6	74	48	ND	Tli pol	53
(TATGCATA)6	74	48	ND	Tli pol	53
(AGATATCA)6	74	48	ND	Tli pol	53
(AGATTCT)6	74	48	ND	Tli pol	53
(AGGGCCCT)6	74	48	ND	Tli pol	53
(GC)24	74	48	ND	Tli pol	53
(TAGG)10/(CCTA)10	70	40	10000	Taq pol	54
(TTAAGGGG)5/(CCCCTTAA)5	75	40	10000	Taq pol	54
(TTAGGG)5/(CCCTAA)5	75	30	10000	Taq pol	54
(TTGGAG)5/(CTCCAA)5	75	30	10000	Taq pol	54
(TGTGAG)5/(CTCACA)5	75	30	10000	Taq pol	54
(ACGTCA)5/(TGACGT)5	75	30	10000	Taq pol	54

ND: Not determine, bp: base pair

E. coli pol I: Escherichia coli DNA polymerase I, Taq pol: Thermus aquaticus DNA polymerase,

Klenow: Klenow fragment of E. coli pol I, Sequenase: modified T7 DNA polymerase,

HIV-RT: HIV reverse transcriptase, Calf pol α: Calf thymus DNA polymerase α,

mouse pol β: mouse DNA polymerase β, Tth pol: Thermus thermophilus DNA polymerase,

Tli pol: Thermococcus litoralis DNA polymerase

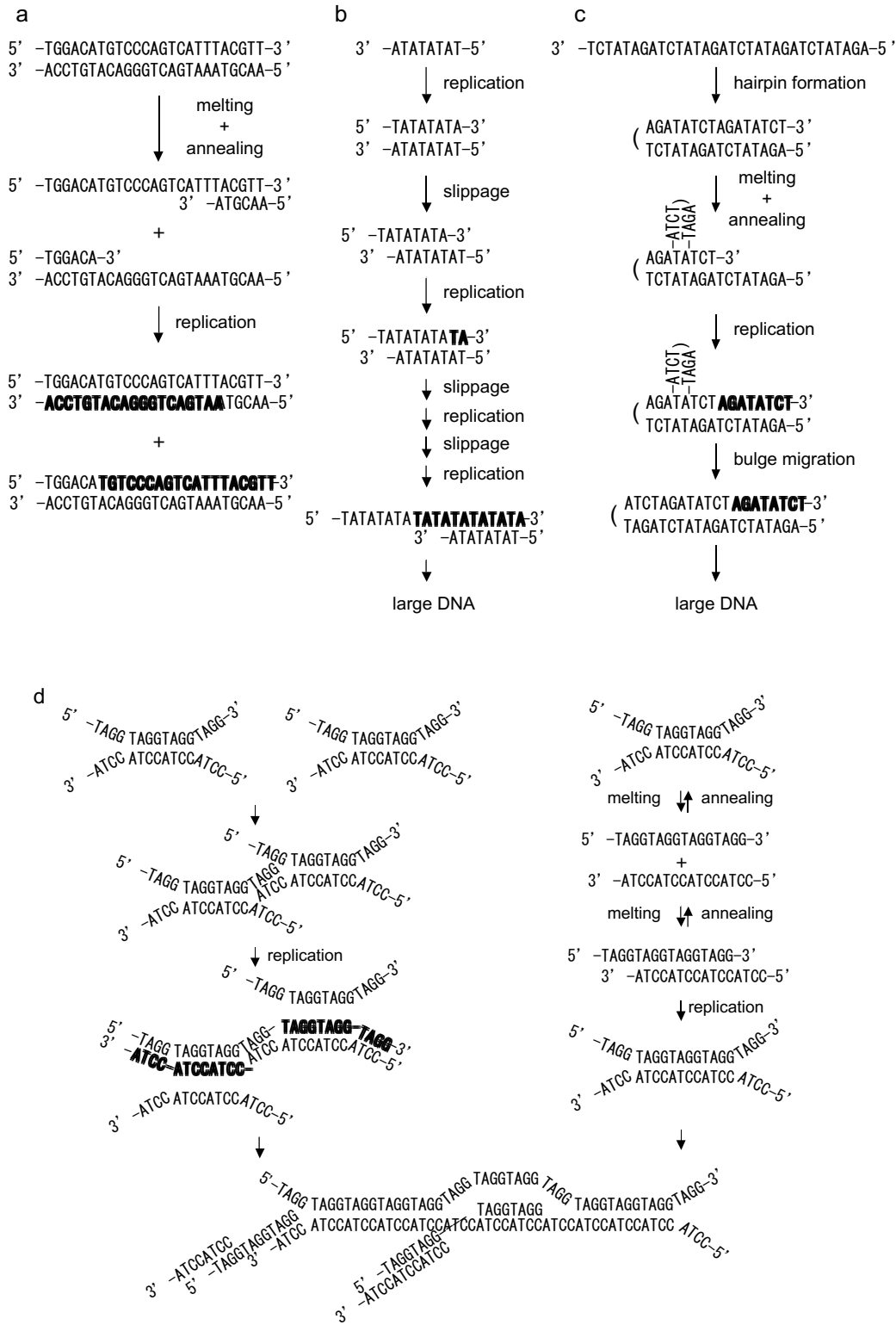


Figure 1. Several models for the repetitive DNA elongation by DNA polymerases
(a) Standard DNA replication mechanism: DNA duplex is melted, and then each single stranded DNA (ssDNA) is annealed with each primer DNA. Complementary DNA is replicated from 3'-OH end of the primer DNA. (b) DNA slippage synthesis model: Newly synthesized stretch of repetitive (TA) DNA is replicated with template (TA) ssDNA. The newly synthesized DNA is "slipped" to the direction of the 3'-OH end of the template. Newly generated ssDNA is replicated by DNA polymerase. Next slippage synthesis goes on until large DNA is made. (c) Hairpin elongation model: Short palindrome repetitive DNA forms a hairpin structure by folding back, and then its end melt partially. The 3'-end arm of the hairpin locally forms a small hairpin, and DNA polymerase elongates the DNA from the 3'-OH end. The small hairpin is next incorporated into the large hairpin by slippage. These cycles of hairpin formation, elongation and slippage are repeated until large DNA is made. (d) Template switching and strand displacement model: Fraying at the ends of duplexes can occur. Due to end fraying, there is transition intermolecular base pairing between duplexes in solution. The strand-displacement activity of the polymerase is then responsible for extending the complementary strands of the DNA duplexes in these intermolecular complexes (Left). At reaction temperatures near the melting point of the duplex, repeated melting and staggered reannealing leads to more efficient extension of the two DNA strands (Right). As the DNA increases in length, bulges form in the duplex at elevated temperature. SsDNA in bulged regions serves as templates for the extension of other duplexes.

なお、PCR に使用する *Taq* DNA ポリメラーゼを用いても、短鎖反復 DNA がスリップ合成で伸長することも記載している。この *Taq* DNA ポリメラーゼによる短鎖 DNA から長鎖 DNA への伸長合成は、*E. coli* DNA ポリメラーゼ I とともに、複数のグループが解析をおこなった[48,49,50]。それぞれ、*Taq* DNA ポリメラーゼと Klenow 断片[51]を用いて、 $(AT)_8$ と $(AAT)_3/(ATT)_5$ という 16 および 9 塩基の短鎖 DNA が、それぞれ 20000 塩基対を超える長鎖 DNA まで伸長した。

その後緒方らは、反復単位長が 4 塩基以上の反復 DNA の伸長例として 2 種類の DNA ポリメラーゼによって 8 塩基の反復単位を持つ短鎖 DNA が長鎖 DNA へと伸長することを報告した[52,53]。これらポリメラーゼの内の一方は、*Taq* DNA ポリメラーゼと同属の好熱性真正細菌 *Thermus thermophilus* から単離された *Tth* DNA ポリメラーゼで、*E. coli* DNA ポリメラーゼ I を含むファミリー A に属する酵素である。他方は、超好熱性古細菌 *Thermococcus litoralis* から単離された *Tli* DNA ポリメラーゼで、真核細胞のゲノム DNA を複製する *pol α* を含むファミリー B に属する酵素である。したがって、ポリメラーゼのタイプが異なっても短鎖 DNA から長鎖 DNA への伸長合成がおこるようである。前者の研究では、8 塩基の反復単位長の 6 回繰り返しからなる 48 塩基の回文反復 DNA $(TACATGTA)_6$ が、74°C で 10000 塩基対の $(TACATGTA)_n$ まで伸長することが示された[52]。後者では、さらに短い 24 塩基長の $(AGATATCT)_3$ が *Tli* DNA ポリメラーゼにより 74°C で 20000 塩基対以上の $(AGATATCT)_n$ に伸長することが示された [53]。両方の伸長合成ともに、反復単位の回文配列を 1 塩基崩した擬回文反復配列でも、DNA の合成量は減少するが伸長合成が観察されている。なお、 $(AGATATCT)_3$ や $(TACATGTA)_6$ を含む多くの 8 塩基回文反復 DNA は 74°C ではヘアピン構造とランダムコイル構造の中間状態となっていることが分光光度計による熱解離特性の解析によって示唆された。ここで、従来のスリップ伸長モデルとは違うヘアピン伸長モデルが提唱された(Fig. 1c)。

最近、反復単位長が 4, 6, 8 塩基からなる 30 から 40 塩基対の非回文反復 DNA が、*Taq* DNA ポリメラーゼにより 70°C で 10000 塩基対以上に伸長することが報告された [54]。伸長させた短鎖 DNA は $(TAGG)_5/(CCTA)_5$ や $(TTGGAG)_5/(CTCCAA)_5$,

$(TTAAGGGG)_5/(CCCCTTAA)_5$ などの 2 本鎖を形成する DNA で、ヘアピン構造は実験条件下では形成しないことが確認された。なお、Klenow 断片でも 37°C でこれら短鎖 DNA が約 100 塩基対まで伸長することも確かめられている。注意深く伸長反応機構を解析したところ、スリップ伸長やヘアピン伸長とは異なり、鋳型 DNA の乗り換え(template switching)[55]と置き換え(strand displacement)[56]を利用して伸長することが示唆された(Fig. 1d)。この報告は、試験管内での反復 DNA の伸長反応において、現在のゲノム DNA 中での反復配列の増減に関する不均等交叉(unequal crossing over)[57]と根本的に類似した鋳型 DNA の乗り換え現象が観察されたことで大変興味深い。

以上の報告から、短鎖反復 DNA を出発材料として、DNA ポリメラーゼはデオキシリボヌクレオチドモノマーを重合し、最長 20000 塩基長の長鎖反復 DNA へと伸長合成できることがわかった。しかし、デオキシリボヌクレオチドの無生物的合成については、第 2 章でもふれたがかなり困難なようである[17]。無生物的に比較的容易に合成できるリボヌクレオチドとその重合体が出発材料となり、後に合成されたデオキシリボヌクレオチドを用いて DNA が逆転写されれば、核酸の無生物合成と長鎖 DNA 合成とはうまく結びつく。そこで次に、短鎖 RNA から長鎖 DNA への試験管内合成に関して示す。

真核生物は、ゲノム DNA の末端に反復配列を合成するテロメラーゼを持っている[58]。この酵素は、タンパク質のほかに RNA (テロメラーゼ RNA) を有する核酸タンパク質(ribonucleoprotein)で、テロメラーゼ RNA の一部の塩基配列を鋳型として利用し、ゲノム DNA の末端に $(TTGGGG)_n$ や $(TTAGGG)_n$ などの反復配列を逆転写的に重合する逆転写酵素である[59]。なお、テロメラーゼは、生体内ばかりでなく試験管内においても、短鎖 DNA や短鎖 RNA を出発材料として長鎖 DNA を合成できる逆転写活性を有している。

24 塩基反復配列 $(TTGGGG)_4$ をプライマーとして、テロメラーゼが、この短鎖 DNA を 5200 塩基長まで伸長することが明らかとなった[59]。テロメラーゼは、プライマーとして、反復配列以外の 20 塩基長短鎖 DNA $(TTTATTTTTTATAAAAATTA)$ や 24 塩基長のランダム配列を有する短鎖 DNA $(AGCCACTATCGACTACGGATCAT)$ を用いることが

でき、3'末端に(TTGGGG)_nを重合し、元の DNA を伸長することもできる [60]. テロメラーゼのプライマーは、通常、DNA であるが、r(GUUGGG)₃ および r(GGGUUG)₃ の短鎖 RNA をプライマーとして用い、鋳型 RNA を DNA へ逆転写できるという報告もある [61]. このことは、テロメラーゼの反応液中に DNA が存在しなくても、RNA を鋳型としてデオキシリボヌクレオチドから反復 DNA を合成できる活性を有することを示す. この DNA 非存在下での RNA 鋳型に対する反復 DNA 合成の逆転写活性はテロメラーゼ以外にも見つかっており、*Taq* DNA ポリメラーゼや *Tth* DNA ポリメラーゼが 20 塩基長反復 RNA (r(AU)₁₀) を鋳型およびプライマーとして反復 DNA(AT)_n を伸長合成することができる[62].

以上の報告から、短鎖 RNA を出発材料として、テロメラーゼの逆転写活性により反復 DNA が合成できることがわかった. この反復 DNA は、DNA ポリメラーゼにより最長 20000 塩基長の長鎖反復 DNA へと伸長合成できることもわかった. 短鎖 RNA の無生物的合成については、第 2 章で言及したように可能である. したがって、短鎖 RNA プールの中にデオキシリボヌクレオチドモノマーが存在していれば、テロメラーゼや DNA ポリメラーゼ活性を持つタンパク質などにより DNA を合成できる可能性は十分考えられるであろう.

5. ゲノム DNA の起源における反復配列の役割

ここまで、(1)核酸の前生物的合成モデルの分析、(2)短鎖反復核酸伸長による遺伝子誕生モデルの分析、(3)DNA ポリメラーゼによる短鎖反復 DNA の長鎖反復 DNA への試験管内伸長合成の分析、(4)逆転写活性を有するタンパク質による短鎖 RNA から長鎖 DNA 伸長合成の分析をおこなった. そこで本章では、(1)から(4)により導き出された知見を用い、ゲノム DNA の起源における反復配列の役割に関して言及する.

核酸の無生物的合成実験から、短鎖 DNA よりもむしろ短鎖 RNA が前生物的に存在した可能性はかなり高くなっている (第 2 章). 最近の縮合剤を使用しない短鎖ペプチドの無生物的合成実験から、前生物的に短鎖ペプチドも存在していた可能性がかなり濃厚になってきている[63,64]. しかし、現存生物中で最小ゲノム

を有するマイコプラズマ・ゲニタリウム(*Mycoplasma genitalium*)でさえ、58 万塩基対と非常に長いゲノム DNA を持っている[7](第 1 章参照). ここまでの知見を用いて生物のゲノム DNA 誕生を考える上で 2 つの問題点がある. それは、この現存生物のゲノム DNA の長さや核酸の前生物的合成モデルで生成できる核酸の長さにあまりにも大きなギャップがある点と、核酸の前生物的合成モデルでは現存生物のゲノムを形づくる DNA の合成が困難な点である. 最初の問題点に関しては、大野が提唱した短鎖反復 RNA が伸長することで長鎖反復 RNA となり、そこから機能を持つタンパク質の情報となる遺伝子(RNA)が誕生したモデル(第 3 章参照)が非常に魅力的である. しかし、これはあくまでもモデルにすぎず、実際の試験管内ではモデルは実証されていない. その上、ゲノム DNA 誕生に必要な、RNA から DNA の変換が説明できていない. そこで筆者は、RNA から DNA を試験管内で合成できるテロメラーゼに注目し、短鎖 RNA から DNA への逆転写的な合成を紹介した (第 4 章). さらに、DNA ポリメラーゼに本来備わっている短鎖反復 DNA を 20000 塩基長以上の長鎖反復 DNA に伸長できる活性に着目し、第 4 章で紹介した. 大野のモデルによれば、反復配列はランダム配列よりも機能を有する遺伝子を生成しやすい. したがって、反復配列が初期のゲノム DNA であった可能性は十分考えられる. これらの知見から筆者は、RNA とタンパク質からなる世界(RNA・タンパク質ワールド)において、第 2 章の前生物的核酸合成モデルにより無生物的に合成された短鎖 RNA が、短鎖ペプチドかまたはリボザイム (触媒活性をもつ RNA) により、短鎖反復 DNA に逆転写的なされ、これが反復 DNA の起源になったのではないかと考えた (Fig. 2). さらに、この短鎖反復 DNA が、短鎖ペプチドかまたはリボザイムの重合活性により、2 万塩基以上の長鎖反復 DNA に伸長し、ゲノム DNA の起源になったのではないかと考えた(Fig. 2). 筆者が提案するこのモデルにおいて、依然、DNA やその部品であるデオキシリボヌクレオチドがどのようにして無生物的に合成されたのかという疑問が残る. 現存生物中でのデオキシリボヌクレオチドはリボヌクレオチド還元酵素のラジカル反応でリボヌクレオチドの 2'-OH を 2'-H に還元することにより生合成される[65]. そこで、デオキシリボヌクレオチドの前生物的合成も短鎖

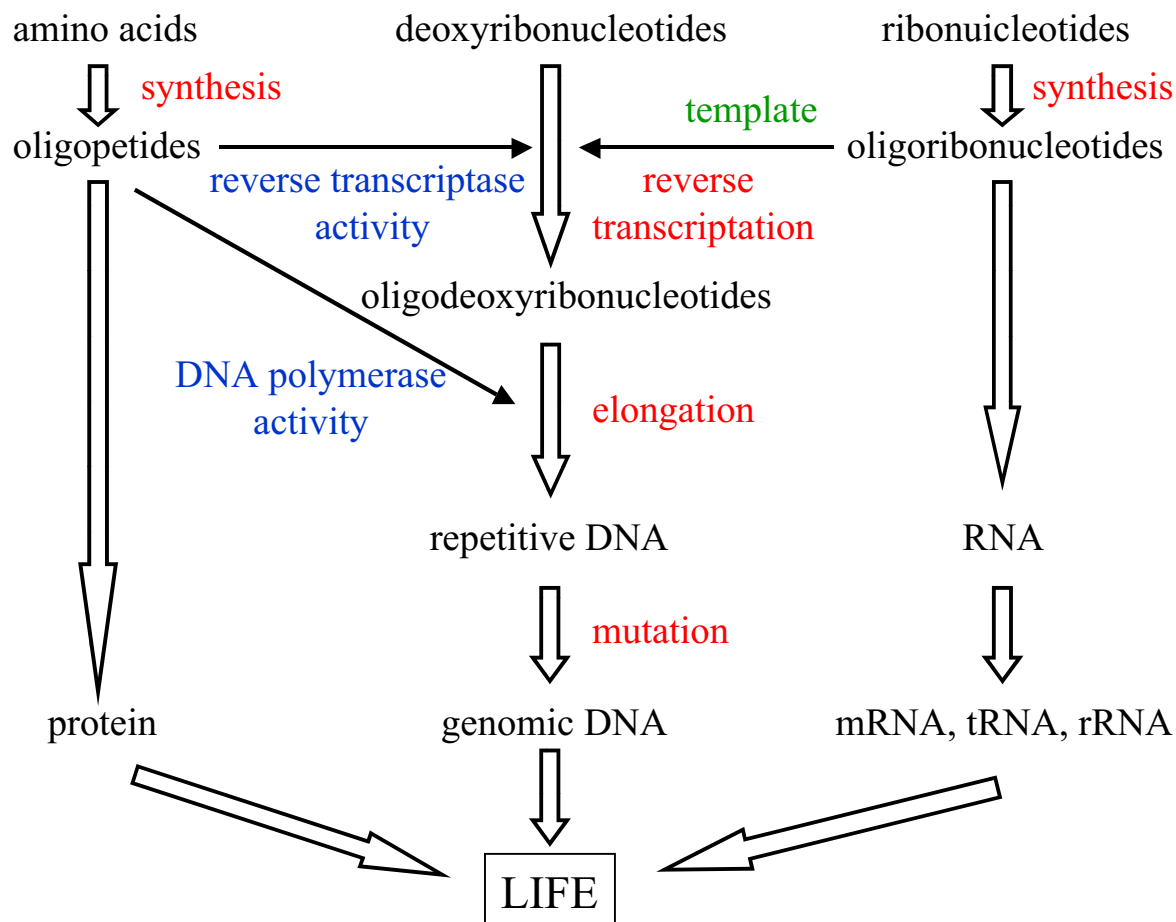


Figure 2. A model for the origin and evolution of genomic DNA

Oligoribonucleotides was synthesized from monoribonucleotides in the presence or absence of catalytic molecules such as a montmorillonite clay under the prebiotic conditions. Some molecules of the oligoribonucleotides synthesized became to the templates of primordial telomerase-like oligopeptides and the catalysts for the synthesis of oligodeoxyribonucleotides (Right). On the other hand, the oligopeptides was synthesized from amino acids under the prebiotic conditions. Some molecules of the oligopeptides synthesized became to the primordial telomerase-like oligopeptides and DNA polymerase-like ones (Left). The catalytic oligoribonucleotides and/or oligopeptides synthesized short repetitive oligodeoxyribonucleotides on the oligoribonucleotide templates by the reverse transcriptional activity. The short repetitive oligodeoxyribonucleotides was elongated to long repetitive DNAs by primordial oligopeptides which have the DNA polymerase-like activity. The long repetitive DNA was evolved to genomic DNA through nucleotide substitution, deletion, and insertion (Center).

ペプチドかまたはリボザイムの還元反応で、リボヌクレオチドから合成したのではないかと考えた(Fig. 2).

最近のゲノムプロジェクトにより、生物のゲノム DNA の塩基配列が多数解析され、ゲノム DNA 中に反復配列が多数存在することが次第に明らかにされてきた。たとえば、ヒトのゲノム DNA には、ゲノム末端に存在する 6 塩基反復配列 (TTAGGG)_n のテロメア[58]、ゲノムの中心付近に存在するセントロメア中の 5 塩基反復配列(GGAAT)_n[66]、およびハンチントン舞踏病などの神経変性疾患の原因となる(CAG)_n/(CTG)_n の 3 塩基単位が反復したトリプレットリピート[67,68]などの反復配列が多数含まれる。また、出芽酵母 (*Saccharomyces cerevisiae*)、ショウジョウバエ (*Drosophila melanogaster*)、および線虫(*Caenorhabditis*

elegans) のゲノム DNA にも、2 塩基反復配列 (GT)_n/(AC)_n[72,73]、(GA)_n/(TC)_n、あるいは(TA)_n[69]や、ゲノム DNA 末端にはテロメア配列として 2 から 4 塩基反復配列(TG₁₋₃)_n (出芽酵母) あるいは 6 塩基反復配列(TTAGGC)_n (線虫) などの反復配列が存在する。真核生物以外の細菌、マイコプラズマ・ゲニタリウムのゲノム DNA にも 3 塩基反復配列(TAG)_n、(CTT)_n、(TGT)_n、(CTA)₈、あるいは(TAC)₅ が存在する [70]。このように、真核生物から真正細菌までの幅広い生物において、反復配列が多数存在することがわかった。

これらのゲノム DNA 中に存在する反復配列は、DNA ポリメラーゼのスリップ合成や不均等交叉(unequal crossing over)、遺伝子変換(gene conversion)により高頻度に反復単位数の増減をおこす[57]。DNA のスリップ

合成は第 4 章で説明した通り、短鎖 DNA の伸長合成に対してもっとも有望であると考え。しかし、不均等交叉がおこるために必要な鋳型 DNA の乗り換え反応は、単純な試験管内での短鎖反復 DNA 伸長合成でもみられ、長鎖反復 DNA の伸長へ一役かっていたかもしれない。反復 DNA が伸長した後に遺伝子が誕生し、ゲノム DNA となるために、不均等交叉や遺伝子変換が遺伝子の再構成するあるいは遺伝子かその近傍の配列が突然変異した経路があったかもしれない。生物のゲノム DNA 中には反復配列が多数散在している。線虫においてはゲノム DNA 末端に通常は存在するテロメア配列が、タンパク質の情報をコードしていないイントロン中にも存在し、ゲノム DNA 中に散在している[71]。したがって、これら反復配列の散在化は、ゲノム DNA の起源と何か関連しているかもしれない。

現在までのところ残念ながら、テロメラーゼの逆転写活性や DNA ポリメラーゼ活性を有する短鎖ペプチドや短鎖 RNA の発見には至ってない。しかし、*in vitro* 選択法[72]などの進化学的手法により、この種の活性が発見される可能性は十分あるだろう。なお、反復 DNA から現在のゲノム DNA の多様な塩基配列への進化に関しては残念ながら論ずることができなかった。この問題に関しては、反復 DNA への伸長合成時の合成エラーの追跡やその系統進化的解析により解決できるかもしれない。

6. まとめ

核酸の無生物的合成に関する実験からは、初期に誕生した核酸は短鎖 RNA である可能性が示唆された。RNA ワールドの視点で考えると、この短鎖 RNA が未だ発見されない方法で伸長して長鎖 RNA となり、初期生命体のゲノムとなっていた可能性が考えられる。そこで、大野が提唱した短鎖反復 RNA が無生物的に生成し、それが伸長することで長鎖反復 RNA となり、そこから遺伝子(RNA)が誕生したモデルは、ゲノムの誕生においても十分検討に値する。しかし、現存生物のゲノムは DNA である。ゲノム DNA の誕生を考える上で、RNA から DNA への変換が必要となる。大野のモデルでは長鎖 DNA の無生物的合成は説明できず、ゲノム DNA 誕生においては別なモデルを提唱しなければならない。そこで筆者は、RNA から DNA を試験

管内で合成できるテロメラーゼに注目し、短鎖 RNA から DNA への逆転写活性が RNA から DNA 生成の役割を演じた考えを述べた。さらに、DNA ポリメラーゼにより試験管内で多く実証されている短鎖反復 DNA から長鎖反復 DNA の伸長合成に着目し、長大なゲノム DNA の誕生に至る十分な長さの生成過程と反復配列の増幅による遺伝情報の誕生過程のモデルを提案した。この考えはゲノム DNA 誕生におけるまったく新しい考えである。今後は、単純な RNA やポリペプチドによる逆転写活性や DNA ポリメラーゼ活性の創製やこれら活性分子と短鎖反復 RNA からなる長鎖反復 DNA の合成、長鎖反復 DNA を試験管内で進化させることによる遺伝子 DNA の誕生を検討する必要がある。

(謝辞)

本報の内容に関して、貴重な御意見をいただいた、市立豊中病院 柴田高氏、大幸薬品株式会社 森野博文氏、京都大学原子炉研究所 齊藤剛氏に感謝いたします。本報の修正に関して、多くのご助言をいただいた、編集責任者の川村邦男氏に深く感謝いたします。

参考文献および注釈

1. Patterson, C. "Evolution" Cornell University Press, Ithaca, New York, 1978.
2. Mojzsis, S. J., Arrhenius, G., McKeegan, K. D., Harrison, T. M., Nutman, A. P. and Friend, C.R., Evidence for life on Earth before 3,800 million years ago, *Nature*, 384, 55-59 (1994).
3. Gesteland, R. F. and Atkins, J. F. Ed., *The RNA World*, Cold Spring Harbor Laboratory Press, New York, 1993.
4. Blattner, F. G., Plunkett, G. 3rd, Bloch, C. A., Perna, N. T., Burland, V., Riley, M., Collado-Vides, J., Glasner, J. D., Rode, C. K., Mayhew, G. F., Gregor, J., Davis, N. W., Kirkpatrick, H. A., Goeden, M. A., Rose, D. J., Mau, B. and Shao, Y. The complete genome sequence of *Escherichia coli* K-12, *Science*, 277, 1453-1474 (1997).
5. International Human Genome Sequencing Consortium, Initial sequencing and analysis of the human genome, *Nature*, 409, 860-921, (2001).
6. The Arabidopsis Genome Initiative (Kaul, S., Koo, H. L.,

- Jenkins, J., Rizzo, M., Rooney, T., Tallon, L. J., Feldblyum, T., Nierman, W., Benito, M. -I., Lin, X., Town, C. D., Venter, J. C., Fraser, C. M., Tabata, S., Nakamura, Y., Kaneko, T., Sato, S., Asamizu, E., Kato, T., Kotani, H., Sasamoto, S., Ecker, J. R., Theologis, A., Federspiel, N. A., Palm, C. J., Osborne, B. I., Shinn, P., Conway, A. B., Vysotskaia, V. S., Dewar, K., Conn, L., Lenz, C. A., Kim, C. J., Hansen, N. F., Liu, S. X., Buehler, E., Altafi, H., Sakano, H., Dunn, P., Lam, B., Pham, P. K., Chao, Q., Nguyen, M., Yu, G., Chen, H., Southwick, A., Lee, J. M., Miranda, M., Toriumi, M. J. and Davis R. W.) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*, *Nature*, 408, 796-815 (2000).
7. Fraser, C. M., Gocayne, J. D., White, O., Adams, M. D., Clayton, R. A., Fleischmann, R. D., Bult, C. J., Kerlavage, A. R., Sutton, G., Kelley, J. M., et al. The minimal gene complement of *Mycoplasma genitalium*, *Science*, 270, 397-403 (1995).
 8. Hutchison, C. A., Peterson, S. N., Gill, S. R., Cline, R. T., White, O., Fraser, C. M., Smith, H. O. and Venter, J. C. Global transposon mutagenesis and a minimal *Mycoplasma* genome, *Science*, 286, 2165-2169 (1999).
 9. Bridson, P. K. and Orgel, L. E. Catalysis of accurate poly(C)-directed synthesis of 3'-5'-linked oligoguanylates by Zn^{2+} , *J. Mol. Biol.*, 144, 567-577 (1980).
 10. Sawai, H. and Orgel, L. E. Oligonucleotide synthesis catalyzed by the Zn^{2+} ion, *J. Amer. Chem. Soc.*, 97, 3532-3533 (1975).
 11. Ferris, J. P. and Ertem, G. Oligomerization of ribonucleotides on montmorillonite: Reaction of the 5'-phosphorimidazole of adenosine, *Science*, 257, 1387-1389 (1992).
 12. Ferris, J. P., Hill Jr., A. R., Liu, R. and Orgel, L. E. Synthesis of long prebiotic oligomer on mineral surfaces, *Nature*, 381, 59-61 (1996).
 13. Ohno, S. Evolution from primordial oligomeric repeats to modern coding sequences, *J. Mol. Evol.*, 25, 325-329 (1987).
 14. Reid, C. and Orgel, L. E. Synthesis in sugars in potentially prebiotic conditions, *Nature*, 216, 455 (1967).
 15. Oró, J. and Kimball, A. P. *Biochem. Biophys. Res. Commun.*, 2, 407-412 (1960).
 16. Reid, C, Orgel, L. E. and Ponnampereuma, C. Nucleoside synthesis under potentially prebiotic conditions, *Nature*, 216, 936 (1967).
 17. Lohrmann, R. and Orgel, L. E. Urea-inorganic phosphate mixtures as prebiotic phosphorylating agents, *Science*, 171, 490-494 (1971).
 18. Osterberg, R., Orgel, L. E. and Lohrmann, R. Further studies of urea-catalyzed phosphorylation reactions, *J. Mol. Evol.*, 2, 231-234 (1973).
 19. Lohrmann, R. Formation of nucleoside 5'-phosphoramides under potentially prebiological conditions, *J. Mol. Evol.*, 10, 137-154 (1977).
 20. Verlander, M. S., Lohrmann, R. and Orgel, L. E., Catalysts for the self-polymerization of adenosine cyclic 2', 3'-phosphate, *J. Mol. Evol.*, 2, 303-316 (1973).
 21. Sawai, H. Catalysis of internucleotide bond formation by divalent metal ions, *J. Amer. Chem. Soc.*, 98, 7037-7039 (1976).
 22. Sawai, H. Oligonucleotide formation catalyzed by divalent metal ions: The uniqueness of the ribosyl system, *J. Mol. Evol.*, 27, 181-186 (1988).
 23. Inoue, T. and Orgel, L. E. Oligomerization of (guanosine 5'-phosphor)-2-methylimidazole on poly(C): An RNA polymerase model, *J. Mol. Biol.*, 162, 201-217 (1982).
 24. Joyce, G. F., Inoue, T. and Orgel, L. E. Non-enzymatic template-directed synthesis on RNA random copolymers, Poly(C, U) templates, *J. Mol. Biol.*, 176, 279-306 (1984).
 25. Joyce, G. F. and Orgel, L. E. Non-enzymatic template-directed synthesis on RNA random copolymers, Poly(C, G) templates, *J. Mol. Biol.*, 188, 433-441 (1986).
 26. Kawamura, K. and Ferris, J. P. Kinetic and mechanistic analysis of dinucleotide and oligonucleotide formation from the 5'-phosphorimidazole of adenosine on Na^+ -montmorillonite, *J. Am. Chem. Soc.*, 116, 7564-7572 (1994).
 27. Follmann. H. Deoxyribonucleotide synthesis and the emergence of DNA in molecular evolution, *Naturwissenschaften*, 69, 75-81 (1982).
 28. Stephen-Sherwood, E., Odem, D. G. and Oró, J. The prebiotic synthesis of deoxythymidine oligonucleotides, *J. Mol. Evol.*, 3, 323-330 (1974).

29. Sherwood, E., Joshi, A. and Oró, J. Cyanamide mediated syntheses under plausible primitive earth conditions, II. The polymerization of deoxythymidine 5'-triphosphate, *J. Mol. Evol.*, 10, 193-209 (1977).
30. Odem, D., Yamrom, T. and Oró, J. Prebiotic oligodeoxynucleotide synthesis in a cyclic evaporating system at low temperatures, *Adv. Spacc. Res.*, 3, 55-59 (1983).
31. Ohno, S. Original domain for the serum albumin family arose from repeated sequences, *Proc. Natl. Acad. Sci. U.S.A.*, 78, 7657-7661 (1981).
32. Ohno, S., Matsunaga, T., Epplen, J. T., Itakura, K., Wallace, R. B. Identification of the 45-base-long primordial building block of the entire class I major histocompatibility complex antigen gene, *Proc. Natl. Acad. Sci. U.S.A.*, 79, 6342-6346 (1982).
33. Ohno, S. and Yazaki, A. Simple construction of human c-myc gene implicated in B-cell neoplasmas and its relationship with avian v-myc and human lymphokins, *Scand. J. Immunol.*, 18, 373-388 (1983).
34. Yazaki, A. and Ohno, S. The recurrence of 49 base decamers, nonomers and octamers within mouse Ig C_{H1} genes and its primordial building block, *Proc. Natl. Acad. Sci. U.S.A.*, 80, 2338-2340 (1983).
35. Ohno, S. The notion of primordial building blocks in construction of genes and transcriptional and processing errors due to random occurrence of oligonucleotide signal sequences, *Adv. Exp. Med. Biol.*, 190, 627-636 (1985).
36. Ohno, S. Immortal genes, *Trends Genet.*, 1, 196-200 (1985).
37. Ohno, S. Repeats of base oligomers as the primordial coding sequences of the primeval earth and their vestiges in modern genes, *J. Mol. Evol.*, 20, 313-321 (1984).
38. Ohno, S. Early genes that were oligomeric repeats generated a number of divergent domains on their own, *Proc. Natl. Acad. Sci. U.S.A.*, 84, 6486-6490 (1987).
39. 3 塩基で 1 コドンとなる。大野は、300 塩基長の遺伝子断片、つまり 100 アミノ酸残基のコドン断片について、停止コドンを含まない断片となる確率を考えた。まず、ランダム配列の場合、開始のコドンが AUG となる確率は 1/64、それに続くコドンがアミノ酸コドンとなる確率は 61/64 となる。この
- アミノ酸コドンが 99 回繰り返すので、99 アミノ酸残基のコドン断片を生成できる確率は $(61/64)^{99}$ となる。そこで、彼は $(1/64) \times (61/64)^{99} = 1.35 \times 10^{-4}$ がランダム配列から遺伝子を生成できる確率と考えた。次に彼は、9 塩基長の反復単位を有する 300 塩基長の遺伝子断片 (100 アミノ酸残基のコドン断片) 中に停止コドンを含まない断片となる確率を考えた。反復配列の場合、基本反復単位である 9 塩基長の塩基配列に停止コドンが入らなければ、鎖長が伸びても停止コドンが入らない。そこで 9 塩基長の遺伝子断片、つまり 3 アミノ酸残基のコドン断片中に停止コドンを含まない断片となる確率からコドン配列を生成できる確率を求めればよい。彼は、9 塩基反復単位の中で、開始コドンのくる場所は 3 カ所あると考え、 $3 \times 1/64$ を開始コドンとなる確率とした。アミノ酸コドンは、残り 2 カ所となるので $(61/64)^2$ が 9 塩基反復単位の中でのアミノ酸コドンとなる確率である。そこで $(3/64) \times (61/64)^2 = 4.23 \times 10^{-2}$ が 300 塩基長の反復配列から遺伝子を生成できる確率と考えた。
40. Kornberg, A. and Baker, T. A. "DNA replication 2nd. ed.," Freeman, New York, pp12-16, 1992.
41. Lehman, I. R., Bessman, M. J., Simms, E. S. and Kornberg, A. Enzymatic synthesis of deoxyribonucleic acid, I. Preparation of substrates and partial purification of an enzyme from *Escherichia coli*, *J. Biol. Chem.*, 233, 163-170 (1958).
42. Kornberg, A., Bertsch, L. L., Jackson, J. F. and Khorana, H. G. Enzymatic synthesis of deoxyribonucleic acids, XVI: Oligonucleotides as templates and the mechanism of their replication, *Proc. Natl. Acad. Sci. U.S.A.*, 51, 315-323 (1964).
43. 通常の DNA 複製は反応が終了すると両端が平滑末端となりそれ以上合成が進まないが、TA 反復 DNA は配列の性質上 2 塩基ずれても再び相補鎖となるため合成が再開できる。つまり、(1) 添加した TA 反復 DNA が鋳型となり DNA ポリメラーゼが相補鎖を合成する、(2) 相補鎖が 5'末端側に 2 塩基スリップすることで 3'末端側に 2 塩基の突出が形成される、(3) 相補鎖の 3'突出末端を DNA ポリメラーゼが複製する、(4) (2)と(3)を繰り返すことで短鎖 DNA が伸長することができる。

44. Wells, R. D., Ohtsuka, E. and Khorana, H. G. Studies on polynucleotides. L. Synthetic deoxyribopolynucleotides as templates for the DNA polymerase of *Escherichia coli*: A new double-stranded DNA-like polymer containing repeating dinucleotide sequences, *J. Mol. Biol.*, 14, 221-240 (1965).
45. Wells, R. D., Jacob, T. M., Narang, S. A. and Khorana, H. G. Studies on polynucleotides. LXIX. Synthetic deoxyribopolynucleotides as templates for the DNA polymerase of *Escherichia coli*: DNA-like polymers containing repeating trinucleotide sequences, *J. Mol. Biol.*, 27, 237-263 (1967).
46. Wells, R. D., Büchi, H., Kössel, H., Ohtsuka, E. and Khorana, H. G. Studies on polynucleotides. LXIX. Synthetic deoxyribopolynucleotides as templates for the DNA polymerase of *Escherichia coli*: DNA-like polymers containing repeating tetranucleotide sequences, *J. Mol. Biol.*, 27, 265-272 (1967).
47. Schlötterer, C. and Tautz, D. Slippage synthesis of simple sequence DNA, *Nucleic Acids Res.*, 20, 211-215 (1992).
48. Ji, J., Clegg, N. J., Peterson, K. R., Jackson, A. L., Laird, C. D. and Loeb, L. A. *In vitro* expansion of GGC:GCC repeats: identification of the preferred strand of expansion, *Nucleic Acids Res.*, 24, 2835-2840 (1996).
49. Hanaki, K., Odawara, T., Nakajima, N., Shimizu, Y. K., Nozaki, C., Mizuno, K., Muramatsu, T., Kuchino, Y. and Yoshikura, H. Two different reactions involved in the primer/template-independent polymerization of dATP and dTTP by *Taq* DNA polymerase, *Biochem. Biophys. Res. Comm.*, 244, 210-219 (1998).
50. Lyons-Darden, T. and Topal, M. D. Effect of temperature, Mg^{2+} concentration and mismatches on triplet-repeat expansion during DNA replication *in vitro*, *Nucleic Acids Res.*, 27, 2235-2240 (1999).
51. *E. coli* DNA ポリメラーゼ I は 3 種類の酵素活性 (DNA ポリメラーゼ活性, 3'-5'エキソヌクレアーゼ活性, 5'-3'エキソヌクレアーゼ活性) を有している。そのうちの 5'-3'エキソヌクレアーゼ活性断片を取り除いた酵素断片を Klenow 断片とよぶ。つまり, DNA ポリメラーゼ活性と 3'-5'エキソヌクレアーゼ活性を有する酵素断片である。
52. Ogata, N. and Morino, H. Elongation of repetitive DNA by DNA polymerase from a hyperthermophilic bacterium *Thermus thermophilus*, *Nucleic Acids Res.*, 28, 3999-4004 (2000).
53. Ogata, N. and Miura, T. Elongation of tandem repetitive DNA by the DNA polymerase of the hyperthermophilic archaeon *Thermococcus litoralis* at a hairpin-coil transitional state: a model of amplification of a primordial simple DNA sequence, *Biochemistry*, 39, 13993-14001 (2000).
54. Tuntiwechapikul, W. and Salazar, M. Mechanism of *in vitro* expansion of long DNA repeats: Effect of temperature, repeat length, repeat sequence, and DNA polymerases, *Biochemistry*, 41, 854-860 (2002).
55. 類似した配列をもつ 2 種類以上の DNA を DNA ポリメラーゼが個別に複製しているとき, 類似した DNA が近接すると, 複製途中の DNA が近接の鋳型 DNA を利用して複製する現象がおこる。この現象を鋳型DNAの乗り換え(template switching)という。
56. M13 フェージの DNA の様な 1 本鎖環状の DNA に相補的なプライマーをアニーリングさせ, DNA ポリメラーゼにより複製を開始すると, 元のプライマーの 5'末端で複製が停止する。しかし, ポリメラーゼの種類により, プライマーや前に複製した DNA をはがしながら複製が再開する場合がある。これを置き換え(strand displacement)という。
57. Richard, G. -F. and Pâques, F. Mini- and microsatellite expansions: The recombination connection, *EMBO Rep.*, 1, 122-126 (2000).
58. Greider, C. W. Telomeres, telomerase and senescence, *BioEssays*, 12, 363-369 (1990).
59. Greider C. W. Telomerase is processive, *Mol. Cell. Biol.*, 11, 4572-4580 (1991).
60. Wang, H. and Blackburn, E. H. *De novo* telomere addition by *Tetrahymena* telomerase *in vitro*, *EMBO J.*, 16, 866-879 (1997).
61. Collins, K. and Greider, C. W. Utilization of ribonucleotides and RNA primers by *Tetrahymena* telomerase, *EMBO J.*, 14, 5422-5432 (1995).
62. Hanaki, K., Nishihara, T., Odawara, T., Nakajima, N., Yamamoto, K. and Yoshikura, H. RNase A treatment of *Taq* and *Tth* DNA polymerases eliminates

- primer/template-independent poly(dA-dT) synthesis, *BioTechniques*, 31, 734-738 (2001).
63. Hill, A. R., Böhler, C. and Orgel, L. E. Polymerization on the rocks: Negatively-charged α -amino acids, *Orig. Life. Evol. Bios.*, 28, 235-243 (1998).
64. Imai, E., Honda, H., Hatori, K., Brack, A. and Matsuno, K. Elongation of oligopeptides in a simulated submarine hydrothermal system, *Science*, 283, 821-833 (1999).
65. Jordan, A. and Reichard, P. Ribonucleotide reductases, *Annu. Rev. Biochem.*, 67, 71-98 (1999).
66. Grady, D. L., Ratliff, R. L., Robinson, D. L., McCanlies, E. C., Meyne, J. and Moyzis, R. K. Highly conserved repetitive DNA sequences are present at human centromeres, *Proc. Natl. Acad. Sci. U.S.A.*, 89, 1695-1699 (1992).
67. Gusella, J. F. and MacDonald, M. E. Trinucleotide instability: a repeating theme in human inherited disorders, *Annu. Rev. Med.*, 47, 201-209 (1996).
68. Warren, S. T. The expanding world of trinucleotide repeats, *Science*, 271, 1374-1375 (1996).
69. Harr, B. and Schlötterer, C. Long microsatellite alleles in *Drosophila melanogaster* have a downward mutation bias and short persistence times, which cause their genome-wide underrepresentation, *Genetics*, 155, 1213-1220 (2000).
70. Hancock, J. H. Simple sequences in a 'minimal' genome, *Nat. Genet.*, 14, 14-15 (1996).
71. The *C. elegans* sequencing consortium. Genome sequence of the Nematode *C. elegans*: A platform for investigating biology, *Science* 282, 2012-2018 (1998).
72. Keefe, A. D. and Szostak, J. W. Functional proteins from a random-sequence library, *Nature*, 410, 715-718 (2001).