

DNA を記憶素子として用いた分子メモリの開発と大容量化に関する検討 Study of Construction of Molecular Memory Based on DNA and Enlargement of Capacity

柏村 聡† 山本 雅人‡ 亀田 充史§ 大内 東‡
Satoshi Kashiwamura Masahito Yamamoto Atsushi Kameda Azuma Ohuchi

1. はじめに

全ての生命現象に係る膨大な情報は、僅か数ミクロンの核中の DNA (Deoxyribonucleic Acid) に A,T,G,C という 4 つの塩基の配列 (塩基配列) としてコード化され保存されていることは良く知られている。また DNA 自体は、化学的・物理的に安定な物質であり、保存された情報を長期に渡って保存し続けることができる。これらを総合すると、DNA は非常に優れた記憶素子としてみることが出来る。近年では、これらの性質に注目し、従来にない超高密度大容量メモリの実現を目指して、DNA を記憶素子としたメモリ (DNA メモリ) の開発が始められている [1],[2]。また DNA メモリは、遺伝情報を保存するメディアや並列アクセスメモリ、生体情報認証を行う場としても期待されている。

DNA メモリは、一般に 1 種類の DNA が 1 つのメモリ単位を表す DNA 集合体であり、各メモリ単位は、その識別子であるアドレス情報とデータ情報の両方が塩基配列として表現されている。DNA メモリへのデータ入力は、入力データとそのアドレス情報を塩基配列にエンコードし、その配列情報を持つ DNA を化学合成し、DNA メモリに追加・混入することで行われる。データの出力は、目的データを保存する DNA のみをアドレス情報に従ってメモリから抽出し、塩基配列の読み取り、デコードすることで行われる。これが、一般的な DNA メモリモデルである。

DNA の化学合成技術が十分発展しているため、メモリへのデータ入力処理に関しては特に問題ない。一方、多種類の DNA 集合から特定の DNA のみを取り出す技術がまだ十分に確立されていないため、DNA メモリ実現の鍵を握るのは読み出し操作の実現であるといえる。従来の DNA メモリは、目的データをコードした DNA を、そのアドレスに対応する配列の相補鎖を結合させることでメモリから取り出している [1],[2]。しかしながら、DNA は相補配列以外にも結合する性質を持っているため、大容量化によって DNA の種類数が多くなると、非目的 DNA が多量に取り出され、目的データのみを読み出しが困難となる。

そこで本研究では、多種類の DNA の中からアドレス情報に対応した DNA のみを正確に取り出すアドレッシング法として「階層的に定義されたアドレスを Nested PCR (Polymerase Chain Reaction) を用いて段階的に指定する」という手法を提案し、それを採用した DNA メモリモデル Nested Primer Molecular Memory (以下 NPMM) を提案した [3]。小規模の NPMM の実現可能性に関しては、化学実験を通して既に示した [3],[4]。本稿では大容量化された NPMM、具体的には 10,000 のアドレス空間を持つ NPMM の実現可能性を検証し、その結果より大容量化することによって生

じる困難な点、問題点を考察する。

2. Nested Primer Molecular Memory

2.1. Architecture

NPMM は、アドレス情報とデータ情報の両方が塩基配列によってコード化された DNA 集合体である点は、通常の DNA メモリと同様であるが、NPMM のアドレス情報は階層的に定義される点が異なる。各階層には複数のユニークな DNA 配列が置かれ、NPMM のアドレス情報は、各階層の各 DNA 配列の組み合わせによって表現される。本稿では、階層数が 4、各階層に 10 種類の配列が用意された NPMM を中心的に扱う (以下 10k-NPMM と呼ぶ)。10k-NPMM のアドレス空間は 10,000 ($=10^4$) である。便宜上、各階層を左から順に A, B, C, D 階層と呼ぶとし、各階層の 10 種類の各配列を $X_i (X \in \{A, B, C, D\}, i \in \{0, 1, 2, \dots, 9\})$ と呼ぶとする (図 1)。Re 部は、後述する PCR という化学反応をアドレス部との間で行うものであり、全てのメモリ単位で共通の配列である。

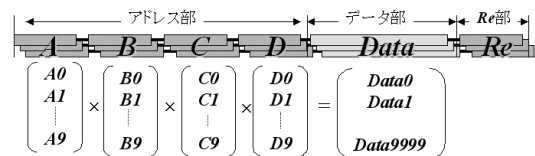


図 1. 10k-NPMM の基本構造

2.2. Addressing operation

NPMM から任意のデータを取り出す操作について説明する。ここでは 10k-NPMM から、アドレス A0B1C2D3 に保存されているデータを取り出す場合を例に取る。これは、 $A_i B_j C_k D_l (i, j, k, l \in \{0, 1, 2, \dots, 9\})$ として定義された 10,000 のアドレス空間を表現する 10,000 種類の DNA 複合体からアドレス情報 A0B1C2D3 を持った DNA のみを抽出することが中心的な操作である。これは以下のように Nested PCR という連続的な PCR によって行われる。まず 10k-NPMM に対して A0 の DNA 配列を持った短い DNA 断片と Re の DNA 断片をプライマーとして用いて PCR という DNA 増幅反応を行う。PCR とは、2 つのプライマー結合領域に挟まれた領域のみを指数関数的に増幅させる性質を持つ酵素反応である。よって 10k-NPMM からアドレス A0BjCkDl ($i, j, k, l \in \{0, 1, 2, \dots, 9\}$) を持つ 1,000 種類の DNA 集合が増幅し、その結果としてメモリ内は A0 を含む DNA 集合でほぼ占められた状態となる。つまり A 階層が A0 である全てのメモリ単位を取り出したことになる。この PCR 反応終了後の状態を希釈した後、B1 と Re をプライマーとして PCR を行くと、A0 と B1 の両方もつ 100 種類の DNA 集合が取り出される。同様に C2 と Re、D3 と Re をプライマーとして用いて連続して PCR を行くと、目的のアドレス A0B1C2D3 をもつ DNA のみが得られる。取り出し

† 北海道大学 大学院 工学研究科

‡ 北海道大学 大学院 情報科学研究科

§ 独立法人 科学技術振興機構

れた DNA の塩基配列を読み取り，デコードすることで情報の取り出しが完了する。

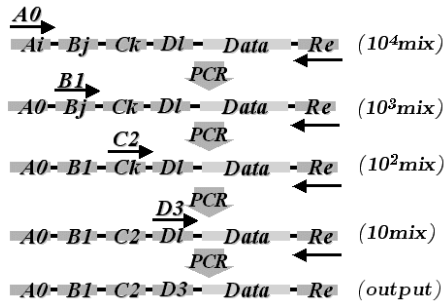


図 2. 10k-NPMM に対して A0B1C2D3 のアドレス指定

2.3. NPMM の特徴

1. エラー排除能：DNA は相補配列以外にも結合する性質を持っている。よって，プライマーが適切な箇所以外に結合した場合，非目的な DNA が増幅される。しかし Nested PCR は，ある段階で生じた非目的 DNA を次の PCR で排除する性質を持った PCR である。よって NPMM はエラー耐性を持つ。
2. 大容量性：階層アドレスのため，少数の DNA 配列により大きなアドレス空間を表現することが可能である。用いる配列数が少ないほど，誤った結合を回避する配列が設計できるため，エラーを抑えつつ大容量化が可能である。
3. 機密性：NPMM 内のデータは，アドレスの各階層を指定するための鍵となるプライマーがなければ取り出すことは出来ない。もしプライマー集合が与えられた場合でも，アドレス情報は順列的に複雑である。よって NPMM は高い機密性を提供する。

3. 化学実験

3.1. 実験概要

本稿では 10k-NPMM を取り扱い，化学実験を通して NPMM の大容量化の可能性を検証する。実験は 2 つのステップに分けられる。第 1 ステップとして，10k-NPMM を化学的に合成する。その後，第 2 ステップとして，合成された 10k-NPMM に対して，幾つかのアドレスを指定して，目的のデータを持つ DNA が取り出されていることを確認する。

10k-NPMM のアドレス空間は 10,000，つまり 10,000 種類のデータが保存可能であるが，本実験では Data0 と Data1 と呼ぶ 2 種類のみを用意することとし， $A_i B_j C_k D_l$ ($i, j, k, l \in \{0, 1\}$) である 16 種類のアドレスには Data0 を置き，それ以外の全てのアドレスには Data1 を置く。敢えて各 Data の頻度にばらつきを与えることで，10k-NPMM が正しく動作するという信頼性を，少ない実験サンプルで得る。というのは，Data0 を保存しているアドレスを指定する操作を行った結果，Data0 が得られた時，10k-NPMM が正しく動作せずにランダムに近い動作をすると仮定すると，その確率は僅か $1/625 (=16/10,000)$ であるからである。

また本実験で扱う 10k-NPMM の各階層間にはリンカと呼ぶ配列を用意する。この配列は構築実験で必要となる。アドレス領域の各配列やリンカ配列等の全ての配列は 20mer と設定し(Data0 のみ 2 本連結した配列 40mer)，10k-NPMM で用いられる全ての配列は，Two-Step Search 法を用いて，互いに誤って結合しないように設計された[5]。Data0 と Data1 の長さを変えることで，電気泳動と呼ばれる DNA の長さに基づいた簡単な実験操作のみで，メモリから抽出されたデータの種類の判別が可能である。

3.2. 構築実験

10k-NPMM は 10,000 種類の DNA 集合である。10,000 種類の DNA を直接合成することは現実的に不可能である。そこで本実験では，リンカ配列が両側に付加された各階層 DNA 配列と，リンカと Re が付加された Data 配列の計 42 種類の DNA を人工合成し，これらを組み合わせることによって 10k-NPMM を化学的に構築した。まず，Data 配列に対して，リンカ付きの D 階層 10 種類の配列と Re をプライマーとして用いて，リンカ配列の部分で結合するように PCR を行った。その結果として，図 3 のように Data の左側に D 階層が付加された 10 種類の生産物が生成された。この操作を A 階層まで繰り返すことで 10k-NPMM を 4 回の化学操作により構築した。またこの時，アドレス $A_i B_j C_k D_l$ ($i, j, k, l \in \{0, 1\}$) には Data0 を，それ以外のアドレスには Data1 を置くように構築した。

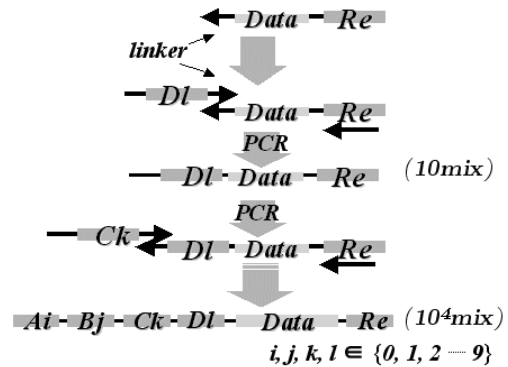


図 3. 10k-NPMM の構築実験概要

3.3. アドレス指定実験

本稿では，表 1 の 7 種類のアドレスをサンプリングし，先ほど合成した 10k-NPMM に対してアドレッシングを行った。10k-NPMM が正しく構築され，意図するように動作すれば，アドレス I, II を指定すると Data0 が得られ，他の 5 つのアドレスは Data1 が得られるはずである。本実験では，Data0 と Data1 に 20mer の差が与えられているため，アドレス指定終了後に 100bp の生産物が得られていた場合は Data0 が，80bp の場合は Data1 が取り出されたことをそれぞれ示す。

図 4. 図 5. は上記の 7 種類のアドレス指定の結果を示した電気泳動写真である。電気泳動とは，DNA の長さに基づいて分離する手法であり，DNA はバンドとして確認される。このバンドの位置と DNA の長さが対応する。

表 1. アドレスサンプル

ラベル	アドレス情報	保存データ
Address I	A0B0C0D0	Data0
Address II	A1B1C1D1	Data0
Address III	A0B1C0D3	Data1
Address IV	A1B0C2D3	Data1
Address V	A1B2C3D0	Data1
Address VI	A2B3C2D1	Data1
Address VII	A3B2C3D2	Data1

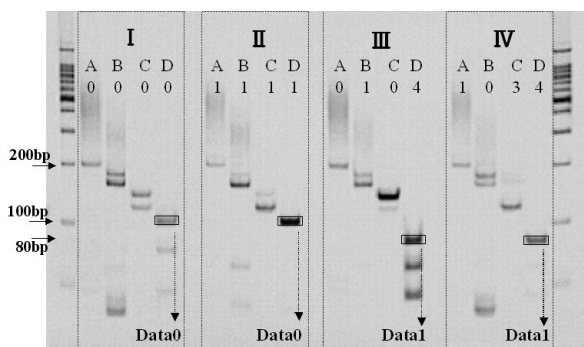


図 4. Address I ~IVの結果

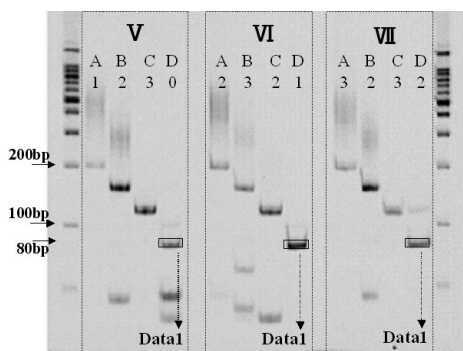


図 5. Address V ~VIIの結果

図 4. の I とラベル付けされた 4 つのレーンが、10k-NPMM に対して Address I をアドレッシングした実験結果を示している。左側の A0 とラベル付けされたレーンは 10k-NPMM に対して、A0 と Re を用いて PCR を行った結果を示し、レーン B0 は、A 階層を A0 に指定した状態に対して B0 と Re で PCR を行った結果を示す。レーン C0, D0 も同様であり、レーン D0 が Address I のアドレッシングが終了した状態を表している。他のアドレスサンプルも同様の表記である。

4. 考察

7 種類全てのアドレスの D 階層を指定したレーンに注目すると、全てのアドレスで目的のデータが得られていることが DNA の長さから確認できる。10k-NPMM が正しく動作しない場合、これらの結果が得られる確率は $(1/625)^2 \times (624/625)^5$ である。つまり 99.99% 以上の信頼性を持って、10k-NPMM は正しく動作しているといえる。また、Address III のレーン C3 のように、アドレス指定処理の過程で、初めて {0,1} 以外の階層値が現れた場合、レーン B1 までは Data0 と Data1 を示す 2 種類のバンドが確認されるが、C 階層を指定すると Data0 を示すバンドが消え

Data1 を示すバンドのみが確認できる。このことから、アドレッシング処理は、正しく目的の DNA のみを取り出し、それ以外を排除していることが確認できる。A 階層のバンドが 1 種類しか見えないのは、Data0 をもつ DNA 集合が Data1 をもつ DNA 集合と比較して圧倒的に少ないため、肉眼でバンドが確認できないだけであると考えられる。また、全てのアドレスの結果において、階層が深くなるにつれ生産物が短くなっていることが確認できるが、これは PCR のプライマーに挟まれた領域のみを増幅する性質によるものである。

Address I のレーン B0 などでは、非目的 DNA が確認されるが、Address I のレーン C0 では非目的 DNA は確認されない。これは、アドレス指定の過程で生じた非目的 DNA が、次のアドレス指定操作の過程で排除されるという前述の NPMM のエラー排除能が働いているからであると考えられる。しかしながら、NPMM をさらに拡張することを考えた場合、非目的 DNA はより生成されやすくなると考えられ、NPMM のエラー排除能のみでは対処できない状況も出てくると危惧される。

また、Address I と Address II の D 階層指定後のバンドを見てみると、明らかにバンドの強度が異なることが分かる。電気泳動の各バンドの強度は DNA 量を反映している。よって、アドレスによっては、読み出しやすいデータ、読み出しにくいデータがあることが確認された。この原因としては、初期状態の各 DNA のばらつきや、アドレス指定の過程で用いるプライマーの反応効率などから生じていると考えられる。さらに NPMM の容量を拡張した場合、このばらつきがより顕著になる可能性があり、その結果として、アクセスすることが非常に困難もしくは不可能なアドレスが生じる可能性がある。

よって、さらなる大容量化を目指すためには、これらを抑えることが重要な課題であると考えられる。

5. おわりに

本研究を通して、10,000 程度のアドレス空間を持つ NPMM ならば実現可能であるということを示した。また同時に、更なる大容量化を目指すための課題も議論することが出来た。

参考文献

1. Baum, E. B. "Building an Associative Memory Vastly Larger Than the Brain", Science, vol.268 (1995)
2. Chen, J., et al, "A DNA based memory with in vitro learning and associative recall", Lecture Notes in Computer Science 2943, (2003)
3. Kashiwamura, S., et al, "Hierarchical DNA Memory based on Nested PCR", Lecture Notes in Computer Science 2568, (2002)
4. Kashiwamura, S., et al, "Potential for enlarging DNA memory: the validity of experimental operations of scaled-up Nested Primer Molecular Memory", BioSystems vol.80 (2005)
5. Kashiwamura, S., et al, "Two-Step Search for DNA Sequence Design", IEICE TRANSACTION, vol.E87-A, No.6 (2004)