

生命の情報

續 順子

<はじめに>

現代社会に大きなインパクトを与えている科学技術の一つとして、バイオテクノロジーを指摘することが出来ます。ひとくりにバイオテクノロジーと言っても、フラスコの中で小さな細胞の固まりを育てることから、細胞の中に特定の遺伝子を運び入れるようなことまで、非常に幅広い技術が含まれます。現在では、生命活動の単位と言える細胞をその素材から組み立てるとまでは行きませんが、ともかく、人間の手によって色々な細胞の性質をブレンドしたり、部分的に変更したりすることの範囲が急速に広がっていることとして、この技術の発展を捉えることが出来ます。私達は、自然界が生み出した材料や素材だけではなく、自分達の手で新しい生命体を生活の中に持ち込み、それらを活用して行く時代を迎えようとしています。

バイオテクノロジーの基礎は、生命活動の設計図、あるいはプログラムとも言える情報が、DNAという物質の構造として書かれていること、そして、今やそのDNAの構造を読み取ること、書き換えることが出来るという点にあります。

私は、将来の食材料の分析や開発研究の基礎技術として、このDNAの構造読み取り、そして更には書き換えをも含めて、ここ数年、DNA構造解析を一つのテーマとして研究を進めてきています。ここでは、今まで取り扱ってきた課題を中心にして、DNAの構造解析の基本的な流れを紹介し、読み取られたDNAの構造から、どのように生命活動の情報が読み出せるのかという所へ話を進めていきたいと思います。

ところで、現在を代表するもう一つの重要な科学技術に、コンピュータ技術があります。情報処理と高度な通信の実現を通じて、物の生産だけでなく知識の生産においても、コンピュータの活用はますます盛んになることでしょう。私達の研究の現場でも、精密な制御の必要な反応機器や、分析機器、そして、大量のデータの保存、処理のためにも、コンピュータの利用は欠かせません。けれども、バイオテクノロジーは一方向的にコンピュータ技術に支えられているだけではなく、バイオテクノロジーが解き明かす生命の情報制御の仕組みは、コンピュータ科学にも大きなヒントを与えるのではないかと思います。

<遺伝情報のありか>

人間や動物、昆虫、植物はもちろんのこと、これらに寄生するバクテリアに至るまで、およそ地球上の生き物は、その体を作り上げている細胞に（バクテリアなどは1細胞で1個体ですけれども）DNAを保持しており、それに各々の生命を維持するための情報を書き込んでいます。DNAは通常単独の分子で存在するのではなく、ヒストンと呼ばれる蛋白質

と組み合わせられて、染色体とか染色糸と呼ばれる構造になっていますが、生命の情報を運んでいるのはDNAです。

分裂して新しく増える細胞には、元の細胞のDNAが正確に複製されて、最初と同じ量のDNAが与えられます。細胞の日々刻々の活動は、機能物質である蛋白質の働きによって営まれています。それぞれの蛋白質をそのパーツであるアミノ酸をつないで組み立てるには、細胞のDNAに書き込まれた情報を読み出さなければなりません。ですから、DNAの構造を知ることは、生命活動の設計図を入手することに相当するわけです。

図1に、DNAの分子模型を示しました。既におなじみのものですが、球で示されているのが炭素や水素、酸素、窒素、燐といった原子で、それらが全体としてらせん階段のような構造を作り上げていることが読み取れます。

このような状態のDNA分子を2本鎖のDNAと呼び、階段の1ステップに相当する部分は、ヌクレオチド対と呼ばれ、A-T, T-A, G-C, C-Gと呼ばれる4種類のものがあります。ここで、A, T, G, Cと記したのは、それぞれアデニン、チミン、グアニン、シトシンという塩基をもつヌクレオチドの略称です。対を作るAとT、GとCを互いに相補的なヌクレオチドと呼びます。

DNAの構造を知るには、こうしたヌクレオチド対の並び方を読み取らなければなりません。ヌクレオチド対が相補的な関係を保っているので、片側の鎖のヌクレオチドの並びが読み取れば、自動的に相手側の鎖のヌクレオチド並びも決まることになります。

さて、私達人間の通常の細胞には、各々46本の染色体が存在しています。これらは23本ずつそれぞれ父の精子と母の卵子から受け継いだもので、この23本のDNAを1本につなげたとすると、その全長はおよそ1m程度、これに含まれるヌクレオチド対は30億個と見積られています。この約30億のヌクレオチド対全てを読み取ろうとする「ヒューマンゲノム計画」という研究計画があります。工場のように設備を組み上げ、コンピュータネットワークで制御されたDNA解析専用の研究施設も最近世界各地に整えられ、1ヶ月で数十万個のヌクレオチド対を読み取ることも可能と言われていますが、それでも一個の精子が運ぶ23本分のDNAの読み取りには1万ヶ月(833年)が必要な計算になります。なお一層の技術開発と、世界的な協同作業が不可欠なのです。

そこで、人間のもののように巨大なDNAでなく(人間のDNAが地上最大なのではありません)、より簡単な短いDNAを持つ生物について、その構造を調べ上げることが現実的な課題であると考えられ、実際にいくつものそうした解析計画が既に完了し、あるいは進行しています。

ウイルスは、生命体と呼ぶことには疑問もありますが、生きている細胞に感染してその

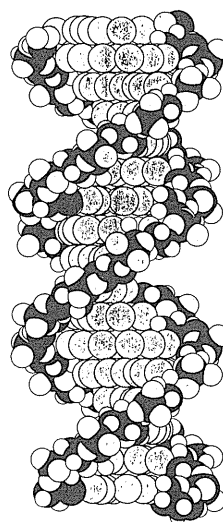


図1 DNAの分子模型
(参考文献1)より
改変)

蛋白質合成能力を乗っ取り、自分と同じものを大量に作らせて増殖します。この指令を与えるものがやはりDNAや類似物質であるRNAです。その大きさは数千ヌクレオチド対程度なので、多くのウイルスについては、その全構造の解析が既に完了しています。HIVウイルスの構造も既に知られていることはご存知でしょう。

細菌（バクテリア）は、数マイクロメートル程度の小さな細胞で、数多くのもが知られています。バイオテクノロジー発展の基礎的実験材料として世界中で利用された大腸菌のDNAは、450万ヌクレオチド対の大きさで見込まれており、もう間もなく読み取りが完成すると期待されています。図2には、様々な生物種のDNAの大きさの比較を示してあります。

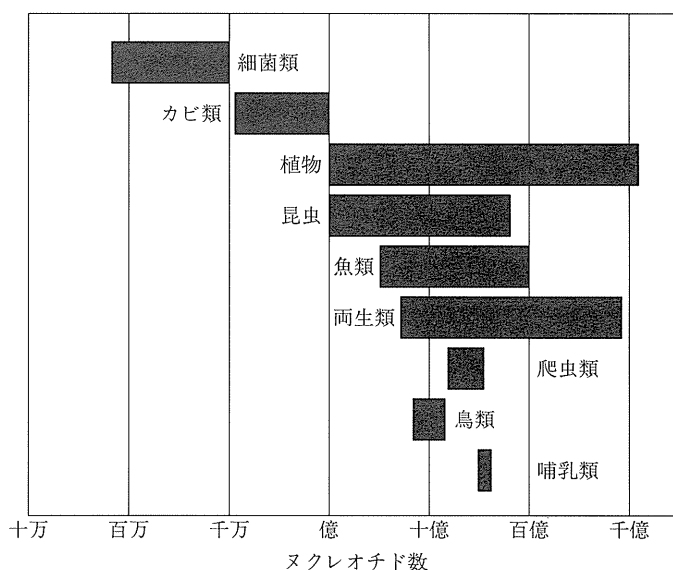


図2 各生物種のDNAの大きさの比較
(参考文献1)より改変)

ところで、動物細胞、植物細胞には呼吸作用を担っているミトコンドリアが、また、植物細胞には光合成作用を行っている葉緑体が含まれていますが、これらはそれぞれの内部に独自のDNAを持っていて、呼吸や光合成に必要な蛋白質の一部を自前で合成しています。構造としての大きさもほぼバクテリアに類似しており、生物の進化の過程で動植物の細胞に取り込まれたバクテリアのようなものが、以来ずっと共同生活を続けて来たものと考えられています。バクテリアとして自活していた時代には、大腸菌に匹敵する大きさのDNAを持っていたのかも知れませんが、動物のミトコンドリアのDNAではほぼ数万ヌクレオチド対、植物のものは一般にこれより大きく、十数万ヌクレオチド対程度、また、葉緑体のDNAは、特殊なものを除いて、やはり十数万ヌクレオチド対程度です。共同生活の間に、核の遺伝子と共通なものが失われたり、一部の遺伝子が核に移ったりして、小さくなっ

たとえられています。

人間の細胞のミトコンドリアが16,569ヌクレオチド対のDNAを持つことが明らかにされたのが1982年、タバコの細胞の葉緑体が155,844ヌクレオチド対であることが明らかにされたのは1986年でした。

DNAの構造のレベルで様々な動植物の違いが広く知られるようになれば、私達を取り巻く生命世界の内容を根底から理解することにつながり、また、そのことを基礎としてより豊かな生命世界の維持発展に進む道が開けることでしょう。バイオテクノロジーに期待されていることは、病気の克服や、環境保全に有用な植物の育成、栄養バランスのよい食物の生産などを初めとして多種多様な内容があります。

私達は、地球上での食物生産の基礎となる光合成作用の場である葉緑体の研究が重要であると考え、タバコの葉緑体DNAに続いてイネの葉緑体DNAの全構造も明らかにしていた名古屋大学の研究グループと共同でクロマツの葉緑体DNAの全構造解析を行いました。クロマツの葉緑体DNAは、119,707ヌクレオチド対で構成されています。葉緑体DNAについては、他に、ゼニゴケ、ミドリムシ、紅藻、トウモロコシなどのものについても全構造決定の報告があり、今後報告予定のものも種々ありますから、葉緑体の遺伝情報の比較検討は、次第に精密なものになって行くことが期待できます。

<DNAの構造解析>

ここでは、クロマツ葉緑体DNAの構造解析を中心に、DNA中のヌクレオチド対の配列（並び方）を読み取る技術的な手続きのあらましを紹介します。

ヌクレオチド配列を読み取るには、目的とするDNAを精製します。植物細胞では、核の中に非常に大きなDNAが何本もあり、その他に葉緑体やミトコンドリアの中にDNAがあります。幸い、葉緑体には同種のDNAが数百本はありますが、それでもこれらの分離精製には労力と時間が必要です。

精製されたDNAから出発しても、最終的にヌクレオチドの一つ一つを読み取る段階では、一度に300ヌクレオチド程度の長さしか読み取れません。ですから、DNAをその大きさの小さな断片に切り分け、しかも、それらを誤りなくつないで元の並びを復元するのに必要な準備が必要です。十数万個のものの並び全体を300個程度の長さに、しかも幾らかのりしろになる重複部分を残して切り揃えると、600種類くらいは最低必要ですし、DNAの2本の鎖それぞれを独立に読み取ることで、読み誤りを無くすようにするのが原則ですから、この二倍以上の種類断片が必要になります。

従って、解析を完成するには、大きなDNAを系統的に断片に切り分ける技術と、ヌクレオチドの配列を読み取る技術、そして、読み取られた配列を全体の構造へ組み立てる技術が組み合わせられるのです。

(DNAの切り分け)

DNAを切り分けるためには、実は大変便利な道具があります。それは、制限酵素と呼ばれる蛋白質で、DNAのヌクレオチド対の特別な部分を探し出して、そこでDNAを切り離し

ます。たとえば、ある制限酵素は、ヌクレオチド対の一方の並びがGGATCC(相手側の並びは、CCTAGG と、丁度裏返しになります) のような部分を認識して、その中の特定の部位を切断します。こうした制限酵素の幾つかを用いて DNA を切り分けることが出来るのです。

切り分けたDNAを分離するには、ゲル電気泳動という技術を用います。これは、寒天のような細かな網目構造をしたゲル状の支持体の中でDNAの断片を電氣的に引っ張ると、DNAは負の電氣を持つので、正の極に引き寄せられますが、網目が邪魔をするので、小さなものは早く、大きなものはゆっくりと移動することを利用して断片を選び分けるものです。図3には、ゲル電気泳動の結果の一例を示しました。

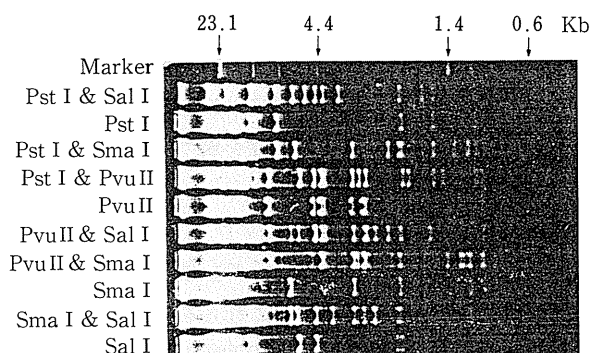


図3 ゲル電気泳動の例
Pst I, Sal I, Sma I, PvuIIは、切断に用いた制限酵素名

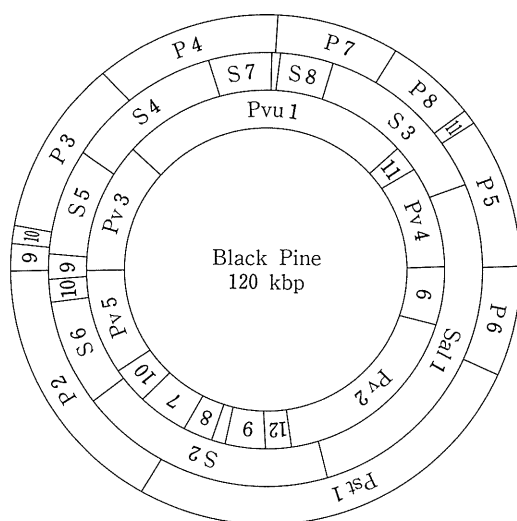


図4 クロマツ葉緑体DNAの制限酵素地図
外側から、Pst I, Sal I, PvuIIによる切断位置をそれぞれ示す。

何種類かの制限酵素でそれぞれ切り分けた結果と、2種類あるいは3種類の制限酵素を同時に作用させて切り分けた結果を総合すると、元のDNAの内部に、どのような順序と距離をおいてそれぞれの制限酵素の認識部位が並んでいたかを、丁度ジグソーパズルを組み立てるようにして決めることが出来ます。このようにして出来上がるそれぞれの制限酵素の認識部位の配置図を、制限酵素地図と呼びます。図4に示したのが、クロマツ葉緑体DNAの制限酵素地図です。クロマツ葉緑体DNAは、既に知られた多くの葉緑体DNAと同様に環状の構造を持っていること、全体の大きさが約12万ヌクレオチド対であることが、この制限酵素地図の作成によって確かめられました。

ところで、これらの制限酵素で切られたDNAの多くは数千ヌクレオチド程度の大きさで、まだまだ直接読み取ることは出来ません。そこで、バイオテクノロジーのもう一つの重要な道具であるプラスミドと呼ばれる特別な環状のDNAを利用します。この環状DNAは、本来は非発病性のウイルスと呼んでよいもので、大腸菌などの細胞に与えると、その中でおとなしく共生して、細胞が増える時に同時に増えて行く性質を持ったものです。これに手を加えて、使いやすい制限酵素の認識部位を幾つか並べて持たせる等の改良が施されています。

葉緑体のDNAを切り分けるのと同じ制限酵素でプラスミドも特定の位置で切り開き、両者を混ぜて、リガーゼというDNAの切り端をつなげる酵素を作用させると、プラスミドが色々な断片を取り込んで再び環状に戻ります。これら大腸菌に与えて培養することで、多様なDNAを必要なだけ増やすことが出来るわけです。このようなプラスミドを集めて、同じ制限酵素で切れば、目的のDNA断片と本来のプラスミドを切り分けられるわけです。千から二千ヌクレオチド対程度に切り分けられた断片を更に切り詰めるのにも、プラスミドを利用します。上で述べた手続きで断片を組み込んだプラスミドを、別の制限酵素で切り開き、そこから長すぎる挿入断片のヌクレオチド対を適当な長さまでにDNA消化酵素をつかって削ってしまうのです。適当な消化時間の後、再び環状につなぐ操作をすれば、このプラスミドをまた大腸菌の中で増やすことが出来ます。消化する時間を調節すると、段階的に短くなった様々な断片を持ったプラスミドが出来ますから、これらをヌクレオチドの個別読み取り操作に持ち込むのです。

(ヌクレオチド配列の読み取り)

DNA断片のヌクレオチドの並び方を読み取る方法として、現在最も一般的なものは、ジデオキシ法と呼ばれるものです。これは、DNA断片の一方の鎖を鋳型にして、それに沿って相手側となるDNAを合成するとき、合成の素材となるA, C, G, Tのヌクレオチドの中に、少量のジデオキシヌクレオチドを加えて反応させるという点が鍵になっています。合成反応が1段階進む毎に、そこで合成を打ち切られたDNAを少量ずつ作らせ、多様なDNA断片を精密なゲル電気泳動で1ヌクレオチドの違いが読み取れる程までに分離し、それぞれの断片がどのヌクレオチドを最後に合成を打ち切ったかを読み取って、ヌクレオチドの並び方を知るのです。

合成反応にも色々細かいトリックが必要ですし、精密なゲル電気泳動を行うにも注意

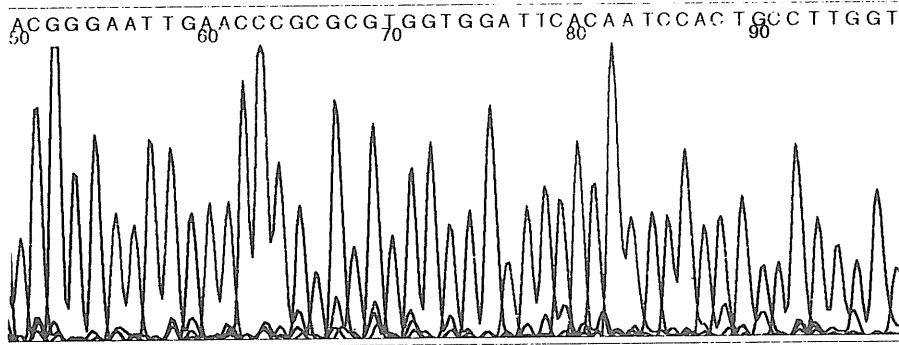


図5 シーケンサーの出力例

深い手続きが必要です。分離されたDNA断片の検出には、現在では主に蛍光を発する物質を利用して、自動化された装置を使います。図5には、自動化された配列読み取り装置の出力の一部を示しておきました。実際の出力は、A, T, G, Cの各々が色分けされて、より読み取りやすくなっています。

実は、こうしたヌクレオチド配列の読み取り操作にも、プラスミドは欠かせない道具です。DNAの合成を開始させるには、目的のDNAとヌクレオチド対を作って結合する20ヌクレオチド程度の短い1本鎖のDNA(プライマー)が必要です。配列の判らないDNAと結合するプライマーを前もって準備することは出来ませんから、実際には、配列を読み出したDNAをプラスミドに組み込んでおいて、プラスミドの一部と結合するプライマーを用いて合成反応を開始させるのです。前節の最後で、様々な長さに切り取られたDNAを含むプラスミドの製法を紹介しましたが、このプラスミドは、そのまま配列読み取りに利用される仕掛けです。

(極く微量のDNAの増幅合成)

少し話の流れが逸れますけれども、ある程度配列が判っているDNAが、極く僅かしか含まれていない試料があるとした場合、例えば、事件の現場に残された血液から、特定部位のDNAを増やしたいような場合、また、化石化した動植物の痕跡からのDNAを増やしたい場合など、次のようにしてその部分を増幅合成することが出来ます。

まず、適切なプライマーを合成します。それを目的の試料に混ぜて、一定時間DNA合成反応を行うと、DNAの複製が起こりますが、このDNAを温めて1本鎖に解離させ、それにまたプライマーを結合させて反応を繰り返すのです。数十回も繰り返せば目的のDNAは元の百万倍、千万倍の量にまで増幅され、解析の作業を進めるのに十分な量となります。

この技術は、PCR法と呼ばれ、繰り返すには厳密な時間間隔での温度管理(加熱と冷却)が必要で、また、その温度変化に耐えて合成反応を行える特殊なDNA合成酵素の発見が不可欠でしたが、プラスミドや大腸菌といった大型の生物的材料を使わず、短時間で急速に目的のDNAを増幅出来るため、DNA解析の応用範囲が大きく広げられることになりま

した。

私達の葉緑体DNA解析でも、上手くのりしろが作れなかった部分の解析を確実にするためや、読み取り結果が不安定な部分の検討のためなどに、このPCR法を用いて目的部分を含む領域を増幅しました。

(断片配列から全体配列へ)

精密な電気泳動を注意深く行っても、読み取れる全ての範囲でヌクレオチドの種類判定が同じ程度に正確であるわけではありません。そこで、少しずつ読み取り位置をずらしながら、また、既に述べたように2本の鎖をそれぞれ別々に読むことによって、データの信頼度を高める努力が必要です。

このため、読み取られる配列データは次々と蓄積し、お互いの間の比較検討作業も膨大な作業量になります。数千ピースの、A, T, G, Cの4文字で書かれた、しかも重複と誤りを含むジグソーパズルを完成させるわけですから、これらを手際良く処理し、データを統一的に管理するには、コンピュータソフトの利用が欠かせません。もちろん、その出発点として、ヌクレオチド配列の読み取りデータ出力をそのまま処理ソフトのデータに変換して、人手による作業の誤り発生を防止することが大切になります。生命の基本情報に迫る道具として、コンピュータはもはや不可欠なのです。

私達が解析したクロマツ葉緑体DNAが、119,707個のヌクレオチド対から成り立っていることも、こうした作業によって確定されました。その全体を書き表すには、少なくともこの雑誌の25ページ以上が必要となりますから、図6には、その一部だけを抜き出して示しました。ここでは、DNAの構造を片方の鎖のヌクレオチド配列で示し、各ヌクレオチドは、A, T, G, Cの一文字で表してあります。

<DNAの保持する情報の解析>

この一見デタラメな記号の羅列が、一体どのような生命の情報を担っているのかと聞かれたら、大抵の人が困惑することでしょう。この羅列をいくら眺めていても、答えを得る

```
ATAGATTAAGTATCTATCCCGTTGGGATAGATATTTAAATCCATCCCAGATATGGTTGACATTGATACATGGATCATATTATACTGTA
AAATAACAAGCCTTATGCTTGGGAGCCTCTGATGATTTTATAAACGAAGTTCTGACCATGACCGCAATTATAGAAAGACGCGAAAGCGCA
AATTTATGGAGTCGCTTCTCGGACTGGATCACTAGCACTGAAAACCGTCTTTACATTGGATGGTTCGGCGTCTTGATGATCCCTACCCTA
TTGACCGCAACCTCTGTATTCAATTATGCTTTTCATCGCAGCTCCTCCAGTAGATATGATGGTATTCGTGAACCTGTTTCTGGTTCTCTT
CTTTATGGAACAATATTATCTCCGGTGCCATTATCTACCTCTGCAGCTATCGGTTTGCACCTTCTATCCCATCTGGGAAGCAGCTTCC
GTCGATGAATGGCTATACAACGGGGTCTTACGAGCTAATCGTTCACACTTCCCTACTTGGTGTAGCTTGTCTATATGGGTCTGAGTGG
GAGCTTAGCTTCCGCTAGGTATGCGTCTTGGATGCTGTGTCATACTCAGCTCCTGTTGCAGCTGCTACTGCCGTTTTCTTGATCTAC
CCATTGGTCAAGGAAGCTTCTGACGGTATGCTCTAGGAATCTCTGGTACTTTCAATTTCAATCAATGATTTCCAGGCTGAGCACAAAT
ATCCTTATGCATCCATCCACATGTTGGGTGTAGCTGGCGTATTCGGCGCTCTCTATTCAAGTATGCATGGTCTTTGGTAACTTCC
AGTTTGATCAGGAAACTACTGAGAATCAGTCCGCAATGCAGGTTAACAATTTGGTCAGGAGGAAGAAACCTACAATATTGTGGCTGCT
CACGGTATTTCCGGCGATGATCTTCCAATATGCTAGTTTCAACAACCTCCCGTCTTTACATTTCTTCTTAGCTGCTGGCCCGTAGCA
GGTATCTGGTTTACCCTCTAGGCATAAGCACTATGGCTTCAACCTAAATGGATTCATTTCAACAATCTGTAGTTGACAGTCAAGGC
CGTGTAATTAACACTTGGGCTGATATCATTAAATCGTGCTAACCTAGGTATGGAAGTTATGCACGAACGTAATGCTCACAACCTTCTCTG
GATCTAGCCGCTGTTGAATCTATTTCAATAGGTGGATAATACTCTGATGGATGTTATCGTGTATTGCTTAATGAAGCATACCAAGCCT
TTCATTCATTTATTGAAAGCTTGGTATGCTTCAATGTATTGTTGGTATTCTTCATACCTTCTCCCATCCATCAATAATGGAT
```

図6 クロマツ葉緑体DNAの塩基配列の一部

ことは難しいと思います。4種類の文字の一つ一つにこだわってはいは問題解決の糸口は見当たりません。

(アミノ酸配列への翻訳)

DNAは、我々のような核のある細胞を持つ生物では、ヒストン・コアと呼ばれる蛋白質に巻き付き、さらに多重に巻き上がった染色体という構造を作って細胞の核の中に格納されています。このDNAのヌクレオチド並びとして保持されている情報は、適切な時期に適切な部分が転写酵素と呼ばれる蛋白質によってRNAの配列として写し取られます。転写されたRNA (mRNA) は核から出て、細胞質と呼ばれる部分にあるリボソームというRNAと蛋白質を成分とする装置に運ばれ、そこでtRNAと呼ばれる小さなRNA分子と出会います。tRNAはその一部に特定のアミノ酸を結合された形でリボソームに付着したmRNAと相互作用し、特定の3ヌクレオチドの組に対してそのアミノ酸をリボソーム表面の合成反応の場に提供するので、これらが次々と連結されて、蛋白質の合成が行われるのです。要約すると、「基本的にはDNAが保持している生命の情報は、RNAが写し取り、RNAの配列をアミノ酸の並びである蛋白質に変換するには3ヌクレオチド単位で翻訳する。」ということになります。

このようなRNAの3ヌクレオチドの組をコドンと呼び、どのコドンがどのアミノ酸に翻訳されるかという関係を表にしたものが、コドン表です。RNAを作るヌクレオチドは、A (アデニン)、G (グアニン)、U (ウリジン、DNAではT:チミンでした)、C (シトシン) の4種類ですから、コドンは $4 \times 4 \times 4 = 64$ 種類が考えられます。アミノ酸は20種類ですので、平均すれば3種類のコドンが同じアミノ酸に対応してよいことになりますが、実際には1種類のアミノ酸に対して、1コドンだけのものが2種、2コドンのものが9種、3コドンのものが1種、4コドンのものが5種、6コドンのものが3種あります。また、UAA, UAG, UGA (DNAでは、TAA, TAG, TGA) の3種類には、対応するアミノ酸がありません。

リボソーム表面でのmRNAとtRNAの作用による翻訳は、特定の位置のAUGコドン(このコドンは、メチオニンというアミノ酸に対するただ1種類のコドンです) から開始され(開始コドン)、アミノ酸に対応するものが無い上述の3種のコドン(終止コドン)のどこかに出会った所で終了することが判っています。

分子生物学の研究の成果として以上のような事情が既に明らかとなっていますので、DNAのヌクレオチド並びを読み取ることができたならば、2本の鎖のそれぞれについてヌクレオチドの並びをコドンに見立てながら読み進み(読む方向はそれぞれの鎖について一定です)、開始コドンに対応したATGの組を発見した位置からは3ヌクレオチド単位に区切って読み進み、終止コドンに出会うまでこれを続ければ、その範囲(ORF)が何かの蛋白質を作るための情報である可能性があります。

この可能性を確かめる手段として現在特に有用なのは、これまでに調査報告された種々の蛋白質のアミノ酸配列との類似性を検討することです。幸い、1980年代から世界的規模でそのような報告をデータベースとして集積し、研究利用のために提供する活動が世界各

地で展開されています。その内容は、最初はテープの形で依頼のあった研究機関などに配布されていましたが、次第にネットワークを利用したアクセスや、後にはネットワークメールによる検索依頼に対してそれぞれのデータベースが結果を報告してくれるような便利なサービスの提供まで行われるようになっていきます。

図7は、図6に示した範囲で組み上げられる最大のORFが、psbAと呼ばれる光合成を行う中心的な蛋白質群の一つであることを、アミノ酸配列の類似性によって示したものです。数種の植物のpsbAについて、配列を比較した結果をまとめてあり、アミノ酸はアルファベット1文字で表わす国際表記に従っています。クロマツ葉緑体のものと同じアミノ酸を持つ場合は、その位置に●を、異なっている場合には文字でそのアミノ酸を表示しています。

(転写過程に付随する情報構造)

図8では、図6で示したヌクレオチドの並びに、図7で確認されたpsbA蛋白質領域を書き入れてあります。ここでは、3個のヌクレオチド並びに対応して、アミノ酸は3文字の記号で表示してあります。ATGからTAAまでの領域が最終的には蛋白質に翻訳される部分で、こうした部分を遺伝子の本体と見なして、それぞれの最終産物に対応した名前を付けるのが一般的です。

ところで、このpsbAの開始コドンの少し上流には、DNAからRNAへの転写に関わる酵素が結合する、プロモーターと呼ばれる部位に特徴的なヌクレオチド並びが認められました。図8に-35、-25、-10と示されているそれぞれ6ヌクレオチド並びの部分がそれです。負の記号がそれぞれの数値に付いているのは、RNAへの転写の開始推定位置より前にある

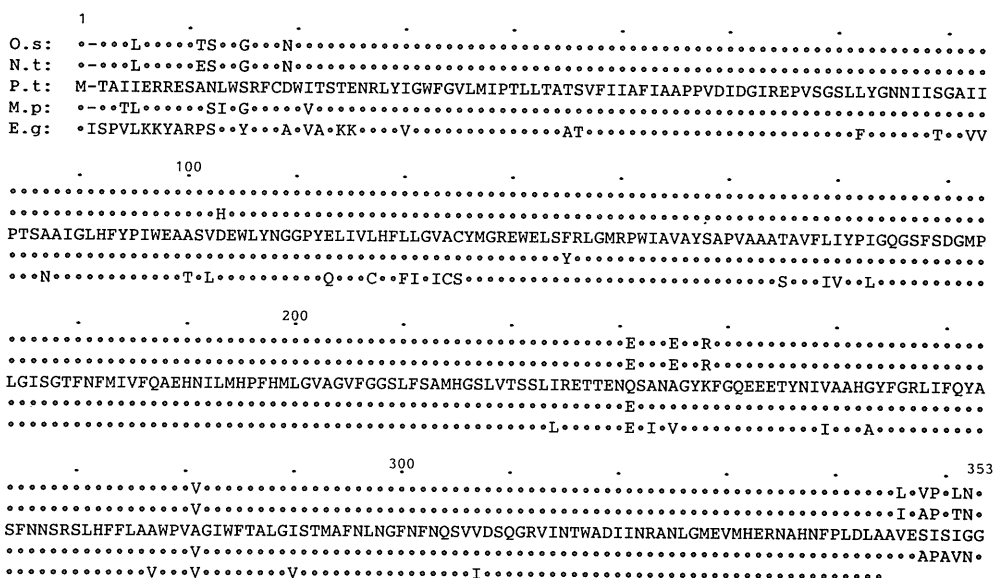


図7 数種の葉緑体DNAのpsbA遺伝子のアミノ酸配列比較
 O.s: イネ、N.t: タバコ、P.t: クロマツ、M.p: ゼニゴケ、E.g: ミドリムシ

-35 -25 -10

ATAGATTTAAAGTATCTATACCCTTCCGATACAAATTTTTAAATTCCATCCCAGATATTGGTTTCACGATTGATTACATCGATCATATTAAATACTGTA
 AAATAACAAGCCCTTATGCTTGGGAGCCTCTGATGATTTTATAAACGAAGTCTGACCATGACCAGCAATTATAGAAGACCGCAAAGCGCA
 MetThrAlaIleIleGluArgArgGluSerAla

AATTTATGGAGTCGCTTCTGCGACTGGATCACTAGCACTGAAAACCGTCTTTACATTGGATGGTTTCGGCGTCTTGATGATCCCTACCCTA
 AsnLeuTrpSerArgPheCysAspTrpIleThrSerThrGluAsnArgLeuTyrIleGlyTrpPheGlyValLeuMetIleProThrLeu

TTGACCCCAACCTCTGTATTTCATTATGCTTTCATCGCAGCTCCTCCAGTAGATATTGATGGTATTCGTGAACCTGTTTCTGGTTCTCTT
 LeuThrAlaThrSerValPheIleIleAlaPheIleAlaAlaProProValAspIleAspGlyIleArgGluProValSerGlySerLeu

CTTTATGGAAACAATATTATCTCCGGTGCATTATTCCTACCTCTGCAGCTATCGGTTTGCACCTCTATCCCATCTGGGAAGCAGCTTCC
 LeuTyrGlyAsnAsnIleIleMetLeuGlyAlaIleIleProThrSerAlaAlaIleGlyLeuHisPheTyrProIleTrpGluAlaAlaSer

GTCGATGAATGGCTATAACAACGGGGTCTTACGAGCTAATCGTTCTACACTTCTACTTGGTGTAGCTTGTATATGGGTCGTGAGTGG
 ValAspGluTrpLeuTyrAsnGlyGlyProTyrGluLeuIleValLeuHisPheLeuLeuGlyValAlaCysTyrMetGlyArgGluTrp

GAGCTTAGCTTCCGTCTAGGATGCGTCTTGGATTGCTGTTGCATACTCAGCTCCTGTTGCAGCTGCTACTGCCGTTTTCTTGATCTAC
 GluLeuSerPheArgLeuGlyMetArgProTrpIleAlaValAlaTyrSerAlaProValAlaAlaAlaThrAlaValPheLeuIleTyr

CCTATTGGTCAAGGAAGCTTCTCTGACGGTATGCCTCTAGGAATCTCTGGTACTTTCAATTTTCATGATTGATTCAGCGCTGAGCACAAT
 ProIleGlyGlnGlySerPheSerAspGlyMetProLeuGlyIleSerGlyThrPheAsnPheMetIleValPheGlnAlaGluHisAsn

ATCCTTATGCATCCATCCACATGTTGGGTGATGCTGGCCTATTTCGGCGGCTCTCTATTTCAGTGCTATGCATGGTCTTTGGTAACTTCC
 IleLeuMetHisProPheHisMetLeuGlyValAlaGlyValPheGlyGlySerLeuPheSerAlaMetHisGlySerLeuValThrSer

AGTTTGCATCAGGAAACTACTGAGAATCAGTCCGCAAATGCAGGTTACAAATTTGGTCAGGAGGAAGAAACCTACAATATTGTGGCTGCT
 SerLeuIleArgGluThrThrGluAsnGlnSerAlaAsnAlaGlyTyrLysPheGlyGlnGluGluGluThrTyrAsnIleValAlaAla

CACGGTATTTCGGCCGATFGATCTTCCAAATATGCTAGTTTCAACAACCTCCCGTTCTTTACATTTCTTCTTAGCTGCTTGGCCCGTAGCA
 HisGlyTyrPheGlyArgLeuIlePheGlnTyrAlaSerPheAsnAsnSerArgSerLeuHisPhePheLeuAlaAlaTrpProValAla

GGTATCTGGTTTACCCTCTAGCATAAGCACTATGGCTTCAACCTAAATGGATTCAATTTCAACCAATCTGTAGTTGACAGTCAAGGC
 GlyIleTrpPheThrAlaLeuGlyIleSerThrMetAlaPheAsnLeuAsnGlyPheAsnPheAsnGlnSerValValAspSerGlnGly

CGTGAATTAACACTTGGCGTGATATCATAATCGTGTAACTTAGGTATGGAAGTTATGCACGAACGTAATGCTCACAACTTCTCTCTG
 ArgValIleAsnThrTrpAlaAspIleIleAsnArgAlaAsnLeuGlyMetGluValMetHisGluArgAsnAlaHisAsnPheProLeu

GATCTAGCCGCTGTTGAATCTATTTCAATAGGTGGATAATACTCTGATGGATTGTTATCGTGTATTGCTTTAATTCGAAGCAATCCCAAGCG
 AspLeuAlaAlaValGluSerIleSerIleGlyGly***

TTTCATTTTATTTAAAGCCCTTCCGATACAAATTTTTAAATTCCATCCCAGATATTGGTTTCACGATTGATTACATCGATCATATTAAATACTGTA

図 8 psbA 遺伝子およびその前後の特異的配列

ことを表すための表示です。

RNAに転写される情報は、遺伝子の本体だけではなく、転写によって生成されるmRNAがリボソームと正しく結合するための手掛かりとなる部分として、ATGコドンの上流にこのような余分とも見られる配列を持っているのです。

さて、図8のpsbA遺伝子本体と推定される配列の下流には、白抜き文字で2組の24ヌクレオチドの並びが強調されています。この部分では、ヌクレオチドが互いに相補的に並んでいますので、ここを転写して作られたRNAは、中間の10ヌクレオチドでUターンして、24ヌクレオチドずつが向かい合って2本鎖の状態になることが出来ます。そうした丁度ヘアピンのようなRNAの構造が出来ることで転写酵素の動きを止め、転写作業を打ち切るきっかけを与えるものと考えられています。つまり、転写されるRNAは、蛋白質に翻訳される部分の末尾にあたるTAAコドンの後ろにも、幾らか余分が付け加わって、全体としては、-35の位置からヘアピンの終りまでと見ることが出来ます。

なお、-35のさらに上流には、15ヌクレオチドの並びと逆向きの相補的な並びが3ヌクレオチドを挟んで向かい合っている箇所が白抜き文字で示してあります。おそらくこれは、この上流にある別の遺伝子に対応したRNAへの転写領域の終端部であると思われます。

慣れない目には乱雑なヌクレオチドの並びとしてしか見えなかった図6の配列が、こうした研究成果を手掛かりにして、まとまりのある、意味をもった構造として捉えられるようになったわけです。DNAの構造解析の目的は、単にDNAを構成するヌクレオチドの並び方を調べるのではなく、こうして、DNAに書きとめられた「意味」を読み取ることなのです。こうした解析を進めるには、多くのDNAの構造を決定して比較解析すると同時に、その意味が現れて行く過程に関与する他の研究の成果を取り込んで、総合的なシステムとして解釈を進める必要があると言えるでしょう。

(アミノ酸配列からの高次元構造予測)

疎水性と親水性：

図8にはpsbA蛋白質を構成するアミノ酸の配列が同時に示されている訳ですが、このアミノ酸配列からは、どのようなより高い次元の情報が読み取れるでしょうか。

生体を構成するアミノ酸には20種類のもものが知られており、人間ではその内8あるいは9種類が必須アミノ酸として栄養的に摂取される必要があります。この20種類のアミノ酸を化学的性質によって分類すると、おおまかには、脂溶性で水になじみにくいもの、水にも脂質にもなじむもの、イオン化して水になじみやすいものに分類出来ます。

psbA蛋白質を構成しているアミノ酸353個の内、216個が脂溶性で、63個がイオン化するもの、残り74個が中間的なものでした。これら各アミノ酸の水とのなじみやすさを評価して、psbA蛋白質全体での分布の様子をグラフとして示したものが図9の折れ線です。図では、縦軸に水とのなじみやすさが目盛りされています。中央から上に出ている部分が水とのなじみの良い部分ということになります。

これを見ると、psbA蛋白質には水になじむものが多い部分と、脂質になじむものが多い部分がある程度まとまって、繰り返し現れていることが読み取れます。脂質になじみやすい部分は、一般に細胞や葉緑体の中にある膜構造の中に潜り込みやすいと考えられ、水になじみやすい部分は、膜から外側へ突き出ていると考えられますから、この蛋白質は大部分膜に埋め込まれているのではないかと想像することが出来ます。別の研究から、psbA蛋白質も含む光合成の中心的役割をする蛋白質のグループが、葉緑体のチラコイドと呼ばれ

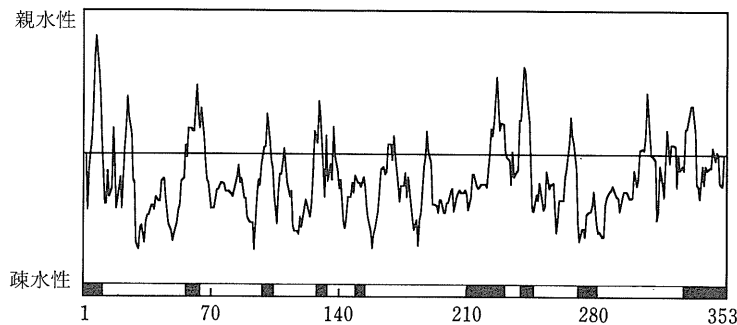


図9 psbA遺伝子のコードする蛋白質の構造予測

る膜に埋め込まれた形で存在することが確かめられており、アミノ酸の種類やこうした計算から予測されることとよく一致しています。

立体構造：

蛋白質を構成するアミノ酸は、単に一列に並ぶのではなく、一部分は折り畳まれたり、螺旋状にねじれたりして、特長のある構造を作ったり、あるいは不規則にまとまっていることが知られています。不規則というのは、デタラメという意味ではなく、それぞれの蛋白質の働きに応じた独特の形ということなのですが、現在のところでは、まだそれぞれの蛋白質のそうした部分の形の間での類似性を表現するためのよい手立てが見当たりませんので、統一的な整理が出来ていないのです。

特長のある構造のうちで、螺旋状の構造を取る条件は、これまでの蛋白質構造の解析の結果からある程度判っていますから、psbA蛋白質のアミノ酸配列について、その条件判定を行ってみたものが図9の下段に示したバーの部分です。黒く塗られている部分が螺旋状構造を取るとされる部分です。こうした部分がいわば蛋白質全体の構造の骨組みになってまとまりを付けていることが想定され、その他のアミノ酸の列が折れ曲がって板状になっている部分などの構造に関する推定の結果を総合すると、psbA遺伝子がコードする蛋白質のおよその形を想像することも出来ます。

精製され結晶化され、詳しくその立体構造が研究された幾つかの蛋白質については、それぞれに特有の反応を引き起こすために必要な構造部分であるアミノ酸配列なども色々と明らかにされていますから、それらと類似の配列を含んでいるかどうかを調べることで、調査している蛋白質の機能について更に詳しい予測を立てることも可能です。

蛋白質を直接取り出してその形を調べるには、精製作業にしても、得られた純粋な蛋白質をX線解析などの方法で調べるにしても、大変な労力と時間、それに設備が必要です。DNAに記述された内容からアミノ酸配列を読み取り、それとこれまでに蓄積された様々な構造研究の成果を比較検討することで、新しく見つかった蛋白質について、ある程度形や機能の見通しが付けられるということは、大変重要なことだと思われま

<生命の情報の読まれ方>

DNAを出発点としたヌクレオチドの配列が一旦アミノ酸の配列に翻訳・変換されて蛋白質の基本構造が決まると、そこでは20種類の多様なアミノ酸の性質を基礎とした新しい構造と機能の関係が現れ、蛋白質は細胞の膜や他の蛋白質、その他の物質と作用して生命活動を支えて行くことになります。こうして、生命での情報の流れでは、基礎的な情報に次第に大きな読み枠を与えて行くことで、色々なレベルでの生命活動の姿に対応して行くことが見えて来ます。

このように、基礎的な情報の単位が次第にまとまりを持ちながら多様な内容を表現するようになるのは、大変一般的な事柄であると思われま

漢字は本来的にはこの単語に相当すると見て良いでしょう。そして、単語の相互配置の規則（文法）を通じて文章を組み立て、更に文章を通じて広い考えを述べあっています。生命の情報表現の世界では、ヌクレオチドが平仮名のような文字に、3ヌクレオチドの組は単語に、そしてアミノ酸の連なりが文章に対応していると考えてよいでしょう。ですから、私達の細胞一つ一つは、それぞれの物語を記述した長い叙事詩を持っていると見ることも出来ると思います。

文章も時に1文字の違い、句読点の打ち誤りによって違った意味になったり、意味が混乱したりします。生命情報の世界では、DNAの1ヌクレオチドの欠落や挿入が、作られるべき蛋白質の構造を変えてしまうこともあります。また、転写開始部位を変化させて、特定の遺伝子の発現の制御が変化してしまうこともあります。こうした生命情報の変化は、様々な遺伝的変異として出現しますし、それらの積み重ねによって、多様な生物種が地上に展開してきたとも考えられます。

バイオテクノロジーの技術は、そうした個々の生命の物語を読み解く手立てを私達に与え、更に、その物語に改変を加える手立てまでも示していると言えます。こうした大きな力を人々の生活の中に正しく活用して行くことは、これからの時代の全ての人間の責任として重たく受けとめられなくてはならないと思われれます。

参考文献

- 1) James D. Watson et. al. "Molecular Biology of the Gene, 4th edition." The benjamin/Cummings Publishing Company, California, 1988
- 2) 杉浦昌弘（編）“クローニングとシークエンス” 農村文化社 1990
- 3) 續順子 梶山女学園研究論集 22号 331-343頁、1991
- 4) Tsudzuki, J. et. al. Mol. Gen. Genet. vol 232,pp 206-214, 1992
- 5) 松原謙一（編）“遺伝子と遺伝の情報1・2” 岩波書店 1992
- 6) 杉浦昌弘（編）“蛋白質・核酸・酵素 植物の分子生物学” 共立出版 1992
- 7) 三浦謹一郎（編）“蛋白質・核酸・酵素 記念特集” 共立出版 1993
- 8) Wakasugi, T., Tsudzuki, J. et. al. Proc. Natl. Acad. Sci. USA. vol 91,pp 9794-9798, 1994