



DNA が作り出す計算

山村研究室 ~ 知能システム専攻



山村 雅幸 助教授

DNAは生体内において遺伝に関わる情報を蓄積している分子である。近年、このDNAそのもので計算を行うDNAコンピュータという研究分野が生まれた。

DNAコンピュータを研究している山村研究室では様々な研究成果や提案を発表している。また、この未知の分野を研究する学生を育成している。DNAコンピュータとはどのようなものか、またどのような用途で使われていくのか、そして山村研究室ではどのような研究をしているのかを紹介していく。



DNA で計算するとは

皆さんは「コンピュータ」という言葉でどのようなものを思い浮かべるだろうか。マウス、キーボード、ディスプレイ、複雑な電気回路……。おそらく多くの人はこのような電子機器を連想するだろう。だが、コンピュータは本当に電子機器でできていなければならないのか。1994年にL.M.エイドルマンが発表したDNAコンピュータは今までのコンピュータとは全く違うものであった。DNAとそれに働く酵素を利用した分子生物学的な操作で数学的な問題を解いたのである。

エイドルマンはDNAをどのように操作し、数学的な問題の解を求めたのだろうか。DNAコンピュータについて説明する前に、DNAの特徴について軽く触れておこう。DNAはA、T、G、Cという4つの要素の連鎖から成る。あるDNA配列に対してAをTに、TをAに、GをCに、CをGに置き換えた配列のことを相補配列といい、相補配列同士は水素結合により結びついて二本鎖になることができる。

この相補配列同士が水素結合するという性質を利用して、水溶液中に大量に溶け込んだDNA配列に対して特定の配列を含むものを全てを取り出す

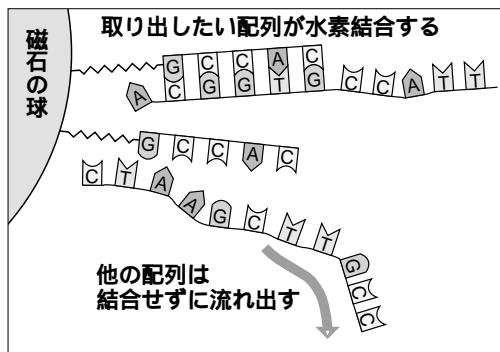


図1 水素結合による条件判断

という操作を行うことができる。まず、望みの配列の相補配列を微小な磁石の球に無数に付けたものを用意する。次にこの磁石の球をDNA配列が溶け込んだ水溶液中に入れると、望みの配列を含むDNAだけが磁石の球についた相補鎖に結合する(図1)。そして磁石で球を引き寄せたまま溶液を流し出すことで、望みの配列を含むDNA配列全てを取り出す。この操作は、溶液中の全てのDNA配列を特定の配列を含むか含まないかで判断をしている。このような判断を条件判断とい

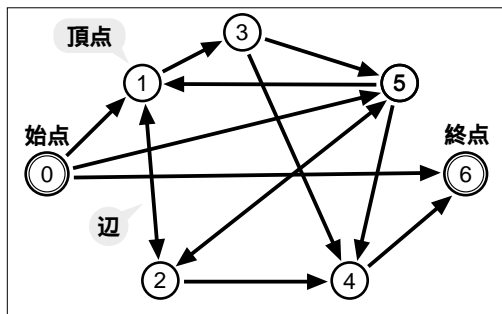


図2 エイドルマンが解いた問題

う。分子生物学で用いられる一般的な実験操作を条件判断と捉え直すことで、エイドルマンはDNAコンピュータを考案したのである。

このDNAコンピュータには、我々が使用している電子コンピュータにはない幾つかの強力な性質が備わっている。まずDNAには高密度に情報を格納できる。例えば、1gのDNAを乾燥させると1cm³ほどの体積になるが、その中に3兆枚のCDと同量の情報を格納できる。

さらに、水溶液1mlあたり10¹²本程度溶け込んだDNA鎖に対して同時に条件判断を行うことが可能だ。先程の例だと、磁石の球に無数に付いたDNA配列は水溶液中のDNA配列をそれぞれ同時に判断し吸着させることができる。また、DNA配列の量が増えても充分な量の磁石の球さえあれば常に一定の時間で操作が終わる。それに対し、電子コンピュータでは沢山のものを同時に計算することはできない。順に一つ一つ計算していくしかないため、計算をするものが増えたと計算時間も増えてしまう。DNAコンピュータは、電子コンピュータでは考えられないほどの並列した計算を可能にしているのである。

エイドルマンが1994年にDNAコンピュータで解いた問題は次の通りである。

「図2の始点から始まり1～5の全ての頂点を經由して終点にたどり着く経路を求めよ。ただし、図2中の矢印は頂点同士を結ぶ辺であり、矢印の方向にしか進めない」

この種の問題は電子コンピュータでは効率的な計算方法が見つからず、問題の規模が大きくなると解を得るためには気の遠くなるほどの時間がかかってしまう。そこで、エイドルマンはDNAを用いた並列計算に着目してこの問題の解を得ようとしたのである。

この問題をDNAコンピュータで解く操作は、

1. 図2から考えられる経路を全て生成する
2. 生成された経路の中から解となるものだけを取り出す

という二つの操作に大きく分かれる。それぞれの操作を詳しく説明しよう。

最初に各頂点をそれぞれ特定のDNA配列で表す。後のために、各頂点のDNA配列の長さを同じにしておく。そして各辺は、辺の始点を表す頂点の配列の後半分と、辺の終点を表す頂点の配列の前半分とを連結させてできるDNA配列の相補配列で表現する。

次に、設計した頂点と辺のDNA配列を用いて、考え得る全ての経路をランダムに生成する。そのためにまず、頂点と辺のDNA配列を大量に作り、リガーゼという酵素と一緒に試験管に入れておく。ここで、先程辺を両端の頂点の相補配列として設計したので、各頂点を表すDNA配列は対応する辺のDNA配列と水素結合することができる。こうして辺と頂点が順々に水素結合していく。長い二本鎖となったDNA配列には切れ目が存在し不安定なのだが、リガーゼはこの切れ目を繋ぐ働きをする。このような反応が繰り返し行われ、試験管の中には図2上の可能な経路がランダムに生成されていく。生成された経路の中には頂点0か

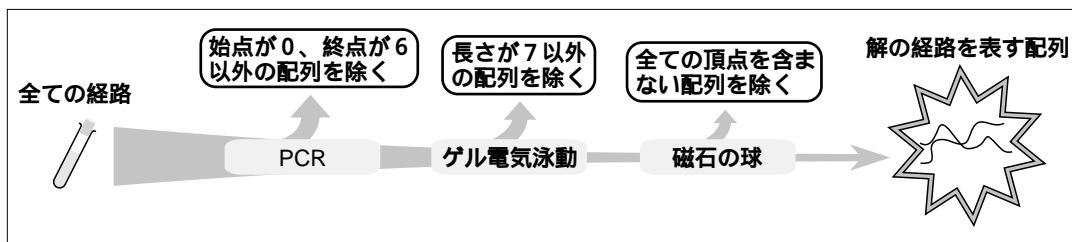


図3 DNAコンピュータによる計算の流れ

ら始まり頂点6で終わるものもあれば、頂点3で始まるものや、頂点5で終わるものもある。試験管の中には、先程も述べたように、1mlあたり 10^{12} 分子程度の膨大な量のDNAが反応している。だから、この反応によってランダムに生成された経路には考えられる全ての経路が含まれているはずだと確率的に考えることができる。そしてそのランダムに生成された大量の経路の中には、解の経路も存在しているはずである。

そこで次に行うべき操作は、**図3**のように試験管の中に含まれるDNA配列をふるいにかけて、解を表すDNA配列のみを取り出す操作だ。

まず、ランダムな経路の中から、解の経路と同じく始点が頂点0、終点が頂点6となっている経路だけを分離する(**図3**)。これにはPCR(ポリメラーゼ連鎖反応)という技術を利用する。PCRとは両端に指定した配列を持つDNA配列だけを大量に複製する技術である。この場合、頂点0を表す配列と頂点6を表す配列を両端に持つDNA配列だけを大量に複製することで、他のDNA配列は相対的にほとんどない状態にする。この操作は**図4**のように複製させたい配列を2倍に増やすサイクルを繰り返す。そうすると始点が頂点0、終点が頂点6である配列のみが指数的な速さで複

製される。それに対し開始頂点が正しいだけの配列と、終了頂点が正しいだけの配列は、ゆっくりとした速さでしか複製されない。そして頂点0も頂点6も端にない配列は一つも複製されない。したがって、十分な時間PCRを行うと試験管内のDNAは頂点0で始まり頂点6で終わる配列が大部分を占めているはずである。

次に、こうして増やした頂点0で始まり頂点6で終わる配列のうち、目的の解と同じ7回だけ頂点を通る経路のみを抽出する(**図3**)。ここで用いるのがゲル電気泳動法である。前の操作で得られたDNA配列の溶液を板状のゲルの端に置いて板の両端に電圧をかける。DNA配列は水溶液中では負の電荷を帯びるため、正の電極に向かって移動していく。この時DNA配列は長さによって移動速度が異なり、移動距離の差が出る。それぞれの頂点を表すDNA配列は同じ長さにそろえたので、移動した距離によって通過する頂点の数を特定できるため、目的の配列だけが抽出できる。

最後に全ての頂点を含む経路を取り出せば、それが解である(**図3**)。すでに、頂点が0で始まり頂点6で終わる経路のみを取り出しているのので、ここでは頂点1~5を全て含む経路を取り出せば良い。この操作はDNAコンピュータによる**図1**の条件判断の例と同じである。PCRの操作により、DNA配列は二本鎖になっているため、高温にして一本鎖にする。そしてまず頂点1の相補配列を無数に付けた磁石の小さな球を入れ、抽出したいDNA配列を結合させる。これに磁石を近づけると、頂点1を含む経路のみが引きつけられ、頂点1を含まない経路は溶液と共に流し出すことができる。同じ操作を2~5の頂点についても行う。こうして最終的に0 5 2 1 3 4 6と0 1 3 5 2 4 6を表すDNA配列が得られ、それが解の経路を表わすものである。また何も得られない時は問題に解がないと分かる。

以上の「計算方法」は一見分子生物学の実験のような操作であるが、それぞれの操作で酵素や相補配列を巧みに使い上手く条件判断を行っていることが分かると思う。そしてこの実験過程によってコンピュータのように計算ができるのである。

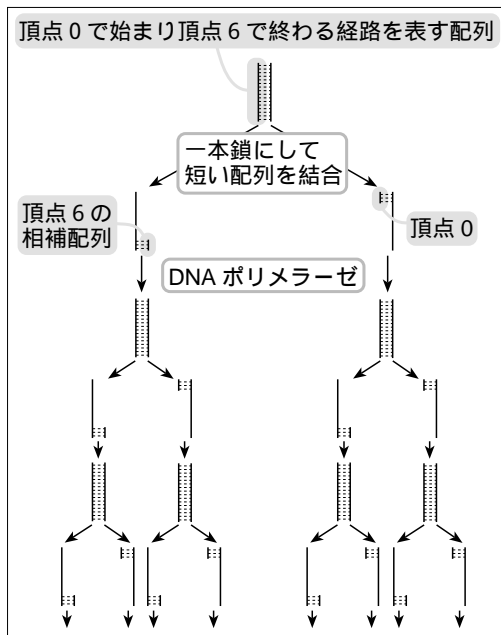


図4 PCR (ポリメラーゼ連鎖反応)



DNA に情報を書き込む分子メモリー

エイドルマンによる問題の解き方は、問題ごとに使うDNA配列を設計し、その組み合わせで解の候補を作り出しそこから適切な解を選ぶというものである。この方法には、問題の規模が大きくなると配列設計のために多くの時間がかかるという大きな欠点がある。それに対して、山村先生が開発した分子メモリーでは情報を書き込むことで解の候補を作るという操作を実現している。これを用いると、問題の設定が少し変わっても情報を書き込む操作を少し変えるだけで対応できる。すなわち、この分子メモリーを用いると配列設計をせずに操作手順だけを考えれば良いのだ。

一般的にメモリーとは情報を記録することのできる装置である。半導体でできたメモリー内の電子素子1つでは0か1しか記録することができない。そこで、複数の電子素子内の0,1の組み合わせでよりたくさんの情報を表現し記録する。この組み合わせられた電子素子による情報を読みとったり記録したりする際に、0または1がどの「場所」の電子素子に入っているかが重要となる。その「場所」をアドレスと呼ぶ。

分子メモリーは、図5-aのような二本鎖で輪になったDNAである。これに特定のDNA配列を認識し切断する働きを持つ制限酵素というものを作

用させる。今回使う制限酵素XはAAGCTTという配列を認識し、DNAを図5-bのように切る。切られたDNAに対して、DNAポリメラーゼを作用させることで図5-cのように全てが二本鎖のDNA配列になる。最後にDNAの接着に使われる酵素、DNAリガーゼで再び輪に戻す。これで書き込みが行われたことになる。実際に輪に戻った図5-dと反応前の図5-aとで比べると、反応後では制限酵素Xが認識するAAGCTTという配列がなくなっていることが分かる。このように分子メモリーでは制限酵素が認識する配列がないかあるかで0,1が記録されているのである。そして制限酵素は様々な種類があり認識する配列が異なるため、それぞれの制限酵素で別々の情報を記録することができる。使う制限酵素をアドレスとし、制限酵素を組み合わせ、たくさんの情報を記録することが可能である。

分子メモリーで解の候補を作るためには、図5のように違う制限酵素で書き込む操作を繰り返す。分子メモリーが入った溶液を分け、別々の溶液に別々の制限酵素で書き込む。分かれた溶液を再び混ぜ合わせ、先程と同じように溶液を分け、別々の制限酵素で書き込む。この操作を様々な制限酵素で行い、それぞれ違う情報が書き込まれた

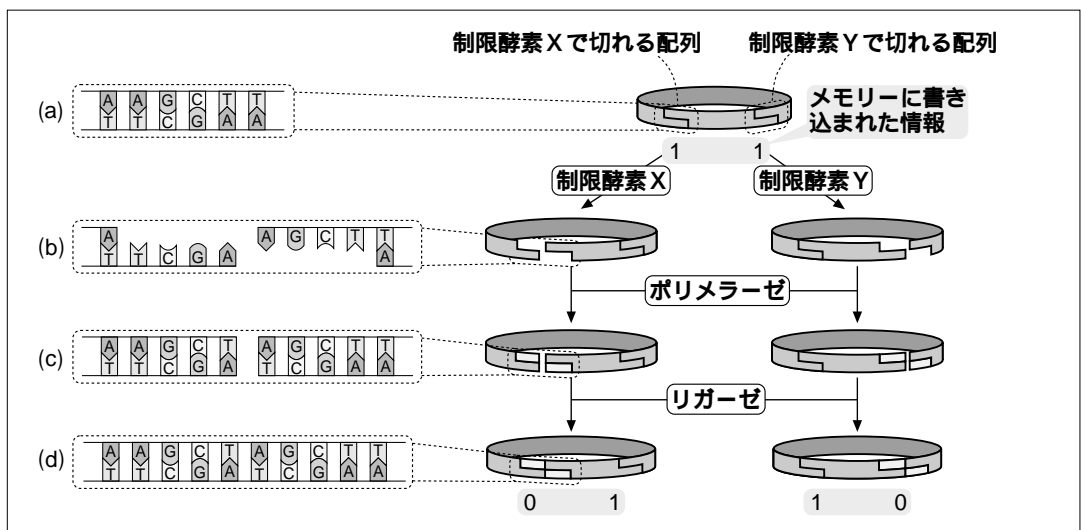


図5 分子メモリー

分子メモリーが解の候補となる。後はエイドルマンが行った問題の解き方と同じように、条件判断を行いつるい分けをして適切な解を求めれば良いのである。

山村先生は十種類の制限酵素を使った十個のアドレスをもつ分子メモリーを開発し、実際に問題を解いた。そして現在は使えるアドレスをより多くしようと、制限酵素を用いず、PNAというDNAによく似ている人工的に生成された分子を用いた分子メモリーも考えている。

分子メモリーも高度な並列性を持つ。大量に存在する分子メモリーに対し、十分な量の酵素さえ

あれば、一つに書き込む時間と同じ時間で全てに書き込むことができる。また書き込まれた回数が一番多い分子メモリーはDNA配列が長くなっているため、ゲル電気泳動法ですぐに取り出すこともできるのだ。

山村先生が開発した分子メモリーはDNAをメモリーとし、特定のアドレスに情報を書き込むというシンプルな機構を確実に実現した。この確実さが重要で、この分子メモリーによって配列設計が一切必要のないDNAコンピュータ開発の基盤を作り出したのである。

DNAによる人工知能

DNAコンピュータは今後どのように応用されていけば良いのだろうか。山村先生はDNAコンピュータで電子コンピュータでも取り扱えるような数学的な問題を解くのではなく、遺伝情報を直接扱うような応用にこそ将来があるのではないかと考えている。それは例えば、体内でガン細胞を認識すると、その細胞を治療するための遺伝子や化学物質を生成するコンピュータである。このような複雑な判断を実現するためにエキスパートシステムという人工知能の一種を使う。

エキスパートシステムとは、すでに分かっている知識（ルール）を大量に組み合わせ、様々な未知の問題に判断をするというシステムだ。特定の分野の知識を蓄えれば、あたかもその分野の専門家のような判断をすることができる。

エキスパートシステムは、電子コンピュータでも過去に盛んに研究されていた。電子コンピュータでは全てのルールを人間がプログラムする。しかし、専門家の知識から抽出してきたルールが全てプログラムできるものではない。ルールには曖昧で定式化できないものもあるのだ。そのため知識を蓄えるのに大きなコストがかかってしまい、エキスパートシステムを作るより専門家を育成するほうが簡単であり、実用にはならなかった。

DNAコンピュータによるエキスパートシステムには様々な利点がある。ルール一つ一つをDNA配列で構成しているため、大量のルールを小さな容積で蓄えられる。また入力に対してそれぞれのルールが条件判断を行う時に、DNAコンピュー

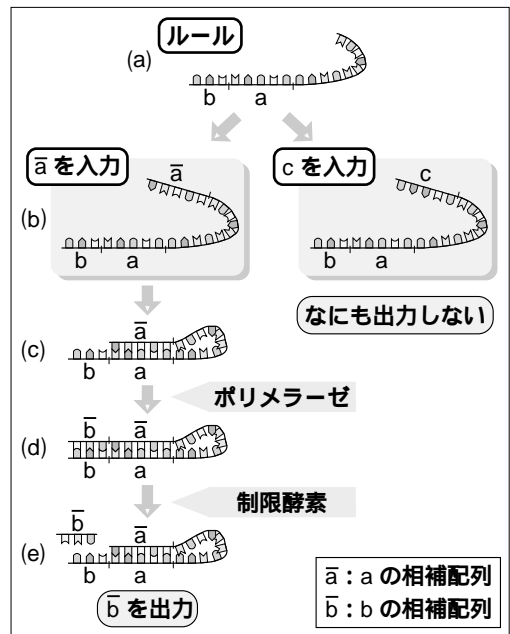


図6 鞭打ちPCRによる条件判断

タの並列性によりシステム全体の判断が高速に行える。そしてポリメラーゼ連鎖反応で、特定の出力をするようなルール全ての濃度を増やしたり、減らしたりするなどが可能である。このため正しい答えが分かっている条件判断を繰り返し行い、その都度正しい答えが出力されるようにルールの濃度を調整することができる。調整をするごとにシステムが今まで解けなかった問題にも、正しい答えを出せるようになっていく。

DNA コンピュータでは一つのルールを、簡単な条件判断を行う鞭打ちPCRで構成する。まず図 6-a のようにルールとなる DNA を用意する。このルールに、入力として a の相補配列を結合させる (図 6-b)。図 6-c のように、この入力の配列とルール部分が結合し、そしてポリメラーゼで相補配列を作り図 6-d になる。そして、特定の配列を認識し切断できる制限酵素を用いることで、b の相補配列を出力することができる (図 6-e)。しかし、ここで a の相補配列以外の入力だとルール部分と水素結合できず出力はない。このルールは a の相補配列が入力されたか、されないかで条件判断をしているのだ。それぞれのルールの DNA 配列の濃度を上げ下げすることで、そのルールに重要度を与えることもできる。これにより、一つのルールでは簡単な条件判断しかできなかった

が、多くのルールを組み合わせることで複雑な条件判断が可能になる。

DNA コンピュータによるエキスパートシステムは生命科学への応用が考えられる。現在、生命科学の分野では、人間の判断で実験データの解釈を行っている。この部分を DNA コンピュータによるエキスパートシステムが条件判断を行い、高速かつ自動化するという応用がある。

分子生物学的実験にコンピュータと同じような性質を見出すことで DNA コンピュータは考え出された。その後、様々な研究者たちが研究を進め発展している。この DNA コンピュータは DNA を扱い研究する分野に様々な形で応用されていくだろう。DNA コンピュータが、生命という大きな分野に新しい考え方を作り出したと言える。

DNA コンピュータの研究をするにあたっては、DNA に関する分子生物学的な知識と計算に関する情報科学的な知識の両方がなくてはならない。どちらか一方の分野の知識だけでは研究できない学際的な分野である。山村先生自身、DNA コンピュータの研究をする以前は人工知能の研究者であった。先生が生物と情報を融合した研究をしようとした時、それまで生物の分野の研究をしていなかったのが、一年半かけてアメリカの大学で生物学の実験と実験室の作り方を習得した。そしてアメリカで得た知識と今まで研究してきた成果を元に山村研究室を立ち上げた。

山村研究室に所属しようとする学生もまた生物と情報の両方を学んでいるわけではない。生物系

から来た学生にはコンピュータ・プログラムを実際に書かせ、情報系から来た学生には分子生物学の実験をさせている。それぞれの分野を学習するというよりは、体験することで他の分野の考え方・学び方を習得させる。このように、研究室にきた人には生物と情報を混合して教育しているが、山村先生はこれを学科でもやるべきではないかと考える。そして、生物と情報の融合した分野で、今までの人工知能の研究ではできなかった知能の本質に近づけるのではないかと考えている。

現在の DNA コンピュータの研究はちょうど、電子コンピュータにおいて、真空管が開発され人間の能力以上の計算をするために盛んに研究されていたころに似ている。そこから半世紀以上、電子コンピュータは様々な大学、研究所、企業で研究され、現在のような高性能かつ小型化された電子コンピュータが手にはいるまでになった。また、用途も計算をするためだけでなく、現在ではネットワークを介して映像・音楽を楽しむこともできるようになっている。DNA コンピュータも数十年後には、現在で考えられているような応用だけではなく想像もできないような用途で使われているであろう。

最後になりましたが、お忙しい中、取材を快諾して下さいました山村先生と研究室の皆様ありがとうございました。(薩摩 兼士)



山村研究室の実験装置