

## DNA の柔軟性を解析する方法

早稲田大学 教育・総合科学学術院 大学院先進理工学研究科 大山隆、木村元、小平玲菜

### はじめに

DNA は柔軟性という視点で見ると不均一な分子である。柔らかい領域もあれば、硬い領域もある。さらに、異方的 (anisotropic) に柔らかい領域もある。柔らかい領域とは、力を加えた時に曲がりやすい領域のことで、硬い領域とは、その逆の性質を示す領域のことである。そして、異方的に柔らかい領域とは、ある特定の方向には曲がりやすいが、それ以外の方向には曲がりにくい領域を指す。柔軟性は、DNA の物理的特性のひとつであり、ゲノムを機能的に折り畳むために利用されているのではないかと考えられてきた。そして班会議でしばしば報告したように、最近になって、この仮説を実証できる可能性がでてきた。DNA の柔軟性の解析は、今後、重要な研究手法になると予想される。そこで、ここではその方法について解説する。

### 解析法

いくつかのグループが短い二本鎖 DNA の相対的柔軟性を明らかにしている。例えば、Brukner らや Satchwell らは、3 bp の DNA 全 32 種類の相対的柔軟性を実験的に求め (表 1)<sup>1,2)</sup>、Packer らは、4 bp の DNA 全 136 種類の柔軟性をエネルギー計算により求めて数値化した<sup>3)</sup>。一般に DNA の柔軟性は、これらのパラメーターセットを用いて以下の手順で解析される。

まず、対象領域をトリまたはテトラヌクレオチドステップの単位に分けて解析してみる [解析単位のことをウィンドウ (window) とよぶ]。例えば、5'-AGCTTAAGCCGG●●●●-3' という配列をトリヌクレオチドステップのウィンドウで、1 bp ずつずらして [スライディングステップ (sliding step: SS) とよぶ] 解析する場合、AGC、GCT、CTT、TTA、TAA、AAG、AGC、GCC、CCG、CGG、GG●、G●●、●●●といった各分節構造の柔軟性の数値を求める (Brukner らのパラメーターセットを用いる場合、表 1 から読み取る)。そして、それらを位置に対してプロットして全体の特徴や内部領域の特徴などを解析する。この段階で何らかの特徴を見出すことができれば良いが、見出せない場合、ウィンドウ

```

1 ATGAAGAAGCT CTGTGTTCAT CAAGCTTTTG GTTTGGCAAT CTCTAGCTGC AGTGGCTCTA 60
TGCCAAGGTT TCGACTTCTT CTACTTCGTC CAACAGGTGA CTCACGTGCA CCATTTACAT 120
GATACACCAC CATGTAACT CACGCCATCA TGTGTTCAGT AGCTCGTCGT TCGAACGGTT 180
CAACCAAATT GATCGAACAG TTCGGTGGTC TTGATCGAAC TGTTAGGTCA ACTTAGCCGA 240
ACTGCTCGAT TCGGCCGAAG TGTTTATAGT CAAGTAGTGA GTTTTTTCTT TGCATTGACT 300
TGTCTCTCA TTTTGAATGT ATATATATAC ATGCAGTGGC CAGGATCATA CTGTGACACA 360
TCGCAGAGTT GCTGCTATCC TACAACITGG AAGCCGGCTT CGGATTTTGG CATCCATGGG 420
CTCTGGCCTA ACTACAACAG TGGTAGCTAT CCATCGAATT GCGACTCCAG CAACCCCTTC 480
GACCCGTCTC AG
    
```

図 1. *dm-1* 遺伝子の部分塩基配列。第 1 エキソン (1-96) の 5'末端から第 2 エキソン (337-492) の 3'末端までの配列を示した。

表 1. Brukner らの柔軟性パラメーター<sup>1)</sup>

| トリヌクレオチド<br>ステップ | 柔軟性<br>(ln p) <sup>a</sup> |
|------------------|----------------------------|
| AAT/ATT          | -0.280                     |
| AAA/TTT          | -0.274                     |
| CCA/TGG          | -0.246                     |
| AAC/GTT          | -0.205                     |
| ACT/AGT          | -0.183                     |
| CCG/CGG          | -0.136                     |
| ATC/GAT          | -0.110                     |
| AAG/CTT          | -0.081                     |
| CGC/GCG          | -0.077                     |
| AGG/CCT          | -0.057                     |
| GAA/TTC          | -0.037                     |
| ACG/CGT          | -0.033                     |
| ACC/GGT          | -0.032                     |
| GAC/GTC          | -0.013                     |
| CCC/GGG          | -0.012                     |
| ACA/TGT          | -0.006                     |
| CGA/TCG          | -0.003                     |
| GGA/TCC          | 0.013                      |
| CAA/TTG          | 0.015                      |
| AGC/GCT          | 0.017                      |
| GTA/TAC          | 0.025                      |
| AGA/TCT          | 0.027                      |
| CTC/GAG          | 0.031                      |
| CAC/GTG          | 0.040                      |
| TAA/TTA          | 0.068                      |
| GCA/TGC          | 0.076                      |
| CTA/TAG          | 0.090                      |
| GCC/GGC          | 0.107                      |
| ATG/CAT          | 0.134                      |
| CAG/CTG          | 0.175                      |
| ATA/TAT          | 0.182                      |
| TCA/TGA          | 0.194                      |

<sup>a</sup>数値が大きいほど柔らかい。計算式に対数が含まれるため負の数値も現れるが、符号自体に特別な意味はない。この幅を拓いて解析してみる。例えば、ウィンドウを 10 bp にしてトリヌクレオチドステップパラメーターを用いる場合、上記配列の最初の 10 bp の柔軟性は、AGC、GCT、CTT、TTA、TAA、AAG、AGC、GCC、CCG、CGG の柔軟性の平均値として求める。このウィンドウで SS を 1 bp にする場合、次の 10 bp の数値は、GCT、CTT、TTA、TAA、AAG、AGC、GCC、CCG、CGG、GG●の柔軟性から同様に計算する。対象とする領域に何らかの特徴があれば、ウィンドウサイズ (WS) や SS の幅を変化させて解析すれば、その特徴はわりと簡単に検出できる。解析の原理自体は、このように至って単純である。しかし、ゲノムワイドの解析を行った時、何千・何万にも及ぶ配列を解析したりするためには、優れたコンピュータプログラムが必要となる。

実際に解析してみよう。どのような DNA でも良いが、ここでは *Dionaea muscipula* (ハエトリソウ) の S-like リボヌクレアーゼ遺伝子を材料にする。少し本題から離れるが、S-like リボヌクレアーゼ遺伝子は、多くの植物でリン酸飢餓

や傷害に応答して発現する。ところが興味深いことに、食虫植物 *Drosera adelae* (モウセンゴケの一種) では腺毛 (捕虫と消化・吸収を司る器官) 特異的に常時発現していることが明らかになった<sup>4)</sup>。この遺伝子 (*da-1* と命名) の発現

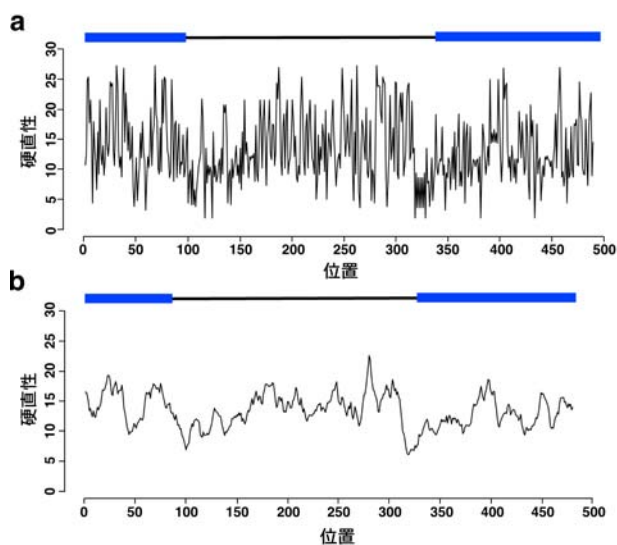


図 2. *dm-1* 遺伝子の部分 DNA 物性。図 1 に示した領域の柔軟性 (硬直性) を Packer らのパラメータ<sup>3)</sup>を用いて解析した。エキソン領域は青線で示した。(a) WS: 4 bp, SS: 1 bp。(b) WS: 10 bp, SS: 3 bp。

制御は、環境応答の視点からも、エピジェネティクスの視点からもたいへん興味深い。我々は最近、*D. muscipula* におけるこの遺伝子のオルソログ (*dm-1* と命名) を明らかにし、現在、その発現制御機構について調べている。

図 1 に *dm-1* の第 1 エキシソンの 5' 末端から第 2 エキシソンの 3' 末端までの塩基配列 (492 bp) を示した。Packer らのパラメータを用いて、この配列の柔軟性を 4 bp の WS と 1 bp の SS で解析した結果が図 2a で、10 bp の WS と 3 bp の SS で解析した結果が図 2b である。彼らのパラメータでは、数値の大きいものほど硬い構造であることを意味する。なお、各数値は、図 2a では各ウィンドウの 5' 側から数えて 2 番目のヌクレオチド番号に対して、図 2b では各ウィンドウの最初のヌクレオチド番号に対してプロットした。これらの図から、領域内の DNA の柔軟性は一定ではないことがすぐわかる。また、第 1 イントロンの両端が柔らかい DNA でできていることもわかる。ここでは示さないが、WS を 25bp にすると、もはやこのイントロンの特徴は検出できない。つまり、WS (と SS) が適切でないと、たとえ柔軟性の特徴が潜んでいても、それを検出することはできない。以上から明らかなように、特徴を“炙り出す”ためには、WS と SS の数値設定が極めて重要である。解析の目的にもよるが、ゲノムワイドの解析では、WS を 1 Mb、SS を 0.1 Mb にすることもある。

#### 参考文献

1. Brukner, I., Sánchez, R., Suck, D. and Pongor, S. (1995) *EMBO J.* **14**, 1812–1818.
2. Satchwell, S.C., Drew, H.R. and Travers, A.A. (1986) *J. Mol. Biol.* **191**, 659–75.
3. Packer, M. J., Dauncey, M. P. and Hunter, C. A. (2000) *J. Mol. Biol.* **295**, 85–103.
4. Okabe, T., Iwakiri, Y., Mori, H., Ogawa, T. and Ohyama, T. (2005) *FEBS Lett.* **579**, 5729–5733.