

GC コンテントの一次元分布情報を用いた一分子 DNA の解析

Single molecular analysis of DNA with one-dimensional distribution of local GC content

杉浦忠男、長屋岳志、佐藤哲大、湊小太郎
奈良先端科学技術大学院大学 情報科学研究科 情報生命科学専攻

Tadao Sugiura, Takeshi Nagaya, Tetsuo Sato, Kotaro Minato
Department of Bioinformatics and Genomics,
Nara Institute of Science and Technology

ゲノム DNA 上では GC コンテントの分布は一様ではない。このことを利用すれば一分子イメージングした DNA を識別できると考えられる。そこで我々はガラス基板上に DNA を伸張固定して GC 特異的に蛍光染色し、エバネッセント顕微鏡を用いて DNA を一分子イメージングすることで DNA の GC コンテントの一次元分布を計測でき、その結果がゲノムデータベースからの予測とよく一致することを確認した。

GC content on genome DNA is not uniform. By use of this feature we can analyze and identify single molecular DNA with one-dimensional distribution of local GC content. We propose a method to analyze single molecular DNA with single molecular imaging by evanescent field microscope and to identify the DNA by comparing measured results of one dimensional distribution of local GC content to estimation from genome database. We have performed the method and have confirmed good correspondence with each other.

1. はじめに

近年、顕微鏡技術の発達と高感度な光検出器の普及によって一分子を対象としたイメージングが多く行われるようになってきている。一つの分子を観測することでこれまでアンサンブル平均としてしか観測できなかった現象をより詳細に分析することができるようになる。対象となる分子もさまざまであるが、この中でも DNA 分子は塩基数で長さを変えられるのでこれまでもさまざまな一分子計測の対象として用いられてきた¹⁻³⁾。

DNA を一分子観測した場合、単に分子を点として観測するだけでなくイメージングすることができるので、塩基配列に関する情報を得られる可能性がある。そこで我々は、局所的な GC コンテントの違いをイメージングすることで一分子 DNA が持つ配列に関する情報を得ることを考案した。生物のゲノム DNA では局所的な GC コンテントがゲノム上で一様でないことはよく知られており、ゲノム解析ではゲノム上の GC コンテントの分布について解析することが行われている。局所的な GC コンテントのばらつきを利用すると一分子 DNA についてもその素性について知ることができると予想される。さらに一分子レベルで DNA の素性が判れば、特定の DNA のみを弁別して回収して解析するなど新たな解析ツールの開発につながることを期待される。また塩

基配列が未知の生物をゲノム解析する際に、ゲノム DNA からさまざまな断片の DNA を作成してそれぞれの GC コンテントを計測し、結果をつなぎ合わせれば GC コンテントの一次元分布を全ゲノム DNA 上で得ることも期待される。さらに GC コンテントの一次元分布はゲノム解析での物理地図として利用することも期待される。

本稿では、一分子 DNA のイメージング法について述べた後、DNA 上の局所的 GC コンテントの分布の計測法について述べる。そして DNA の伸張固定について述べ、実際に伸張固定した DNA のイメージング結果および GC コンテントの一次元分布の計測結果について示す。さらに、GC コンテントの一次元分布をデータベースと照合する手法について述べ、実際に計測結果を照合したので報告する。

2. GC コンテントの一次元分布情報の計測の原理

一分子 DNA をイメージングするためにエバネッセント顕微鏡を用いる。エバネッセント顕微鏡とは、エバネッセント状態になった光を励起光として用いた蛍光顕微鏡で、基板表面近傍のみを蛍光励起して観察できるものである⁴⁾。エバネッセント状態の光とは、光を高屈折率基板側から全反射角以上の角度で入射した場合に低屈折率媒質側に生じる光の場合で、その強度が基板表面から離れるに従って指数関数的に減衰し、その中に蛍光分子が存在すれば蛍光励起する性質を持っている。図 1 にエバネッセント顕微鏡の模式図を示す。レーザーから出た光を顕微鏡対物レンズに入射して試料に対して斜めの方向から入射して照明する。このとき

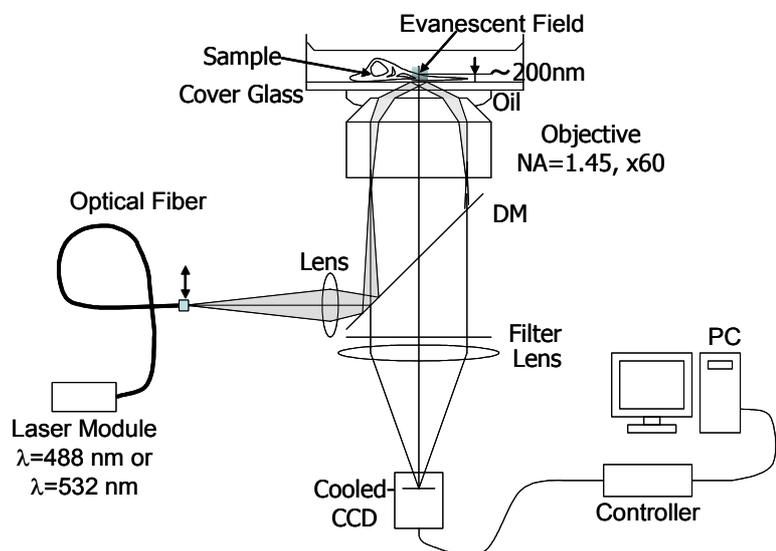


図1 エバネッセント顕微鏡の模式図

の入射角を全反射角以上の角度にすると照明光はカバーガラスの表面で全反射して、カバーガラス表面にはエバネッセント場が形成される。このエバネッセント場は基板から 200 nm 程度の距離までしか到達しない。よってこの照明法を用いることで基板表面に固定された DNA 分子のみを選択的に蛍光励起して観察できるようになる。DNA から発せられた蛍光は同じ対物レンズを通して集められ、冷却 CCD カメラ上に結像する。

DNA は適切な蛍光色素を利用するとほぼ一様に蛍光染色することも GC コンテントの分布に従って染色することも可能である。GC コンテントの一次元分布を計測するために我々は GC 特異的に結合する蛍光分子である 7-aminoactinomycin D (7-AAD) を用いて染色する。また、基板に固定された DNA 位置を特定するためにほぼ一様に DNA を染色する蛍光分子である YOYO-1 iodide を用いて染色する。7-AAD および YOYO-1 による染色はまず 7-AAD による染色をしたあ

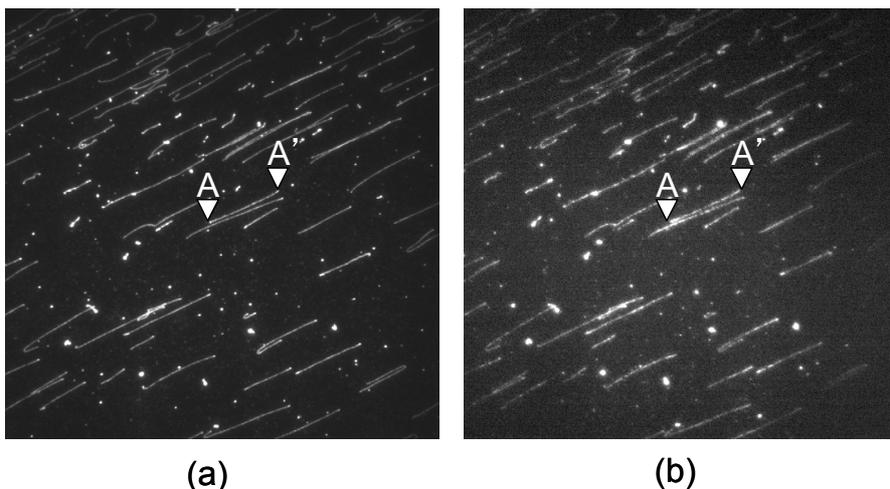


図2 伸張固定した DNA の観測結果 (a)は YOYO-1 の蛍光画像、(b)は 7-AAD の蛍光画像

とで YOYO-1 による染色を行い、2重染色法された状態で DNA を基板上に固定する。これらの蛍光分子は励起波長 λ_{ex} と蛍光波長 λ_{em} が 7-AAD では $\lambda_{ex} = 550 \text{ nm}$ 、 $\lambda_{em} = 650 \text{ nm}$ 、YOYO-1 では $\lambda_{ex} = 490 \text{ nm}$ 、 $\lambda_{em} = 510 \text{ nm}$ であり、励起波長と蛍光波長から用意に識別できる。励起波長、蛍光波長はいずれも発光強度が最大になる波長を示している。

基板上に DNA を固定する際には、DNA は基板上に伸張した状態で固定されているのが望ましい。そこでスペルミジンを用いて DNA を伸張固定する手法を適用する。DNA はスペルミジン水溶液の濃度によってグロビュール状から伸張状態に変化する性質がある。そこで、スペルミジンでグロビュール状にした DNA を基板上に滴下し、その液滴中に純水を注入してピペッティングによって水流を発生させることで濃度を変化させて水流によって DNA を力学的に伸張させて、基板上に伸張状態で固定されるようにする。ガラス表面へ DNA を固定するためにガラスはアミノシラン処理して用いる。

3. 伸張固定した DNA の観測

伸張固定した DNA をエバネッセント顕微鏡で観測した結果の一例を図 2 (a) に示す。DNA は λ DNA で、YOYO-1 で蛍光染色し、波長 488 nm のレーザーで照明して観察した。画像中に複数見られる長い線状の物体がそれぞれ一分子 DNA である。DNA が一本であることは線状構造の切断方向の強度分布がほぼ同じ高さのピークを持つことから確認した。

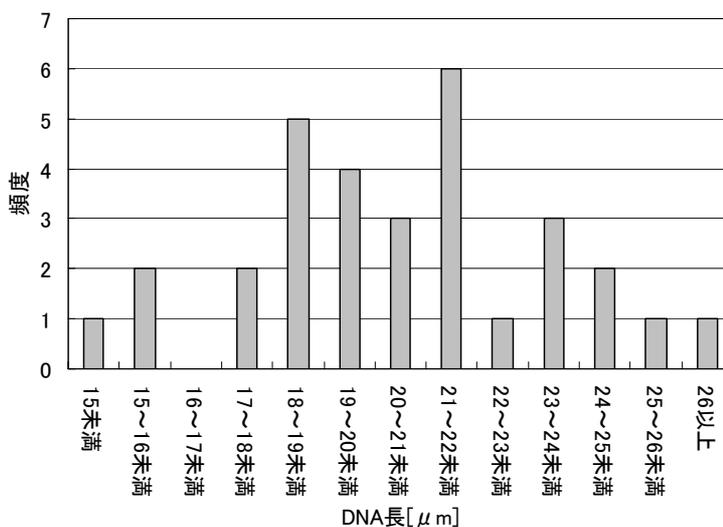


図3 λ DNA の長さの計測結果(ヒストグラム、N=30)

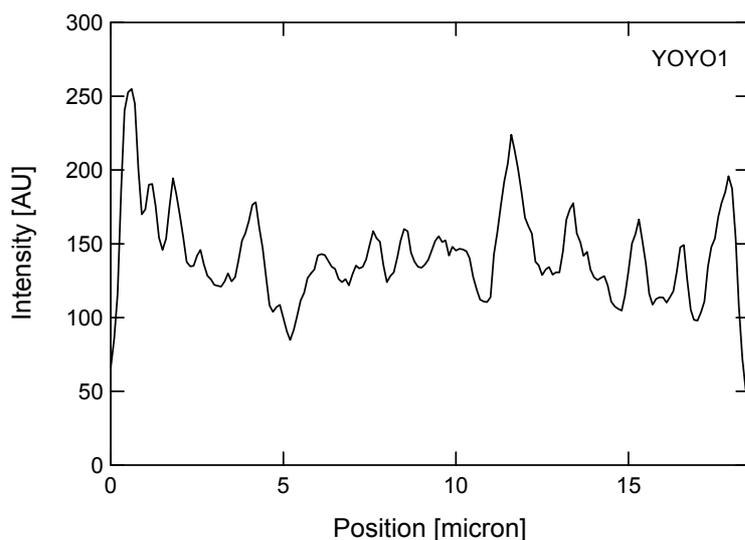


図4 線方向の強度分布 (YOYO-1)

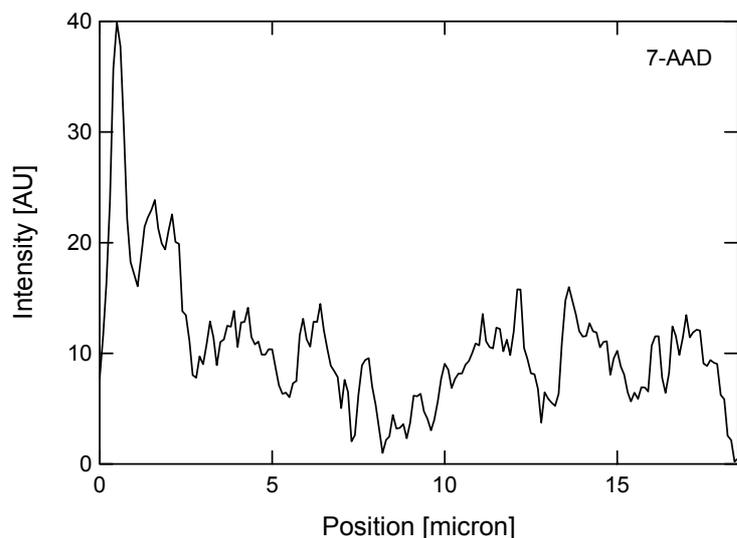


図5 線方向の強度分布 (7-AAD)

それぞれの DNA について解析するために長さの分布を調べた。画像中で DNA 部分を抽出し、線状構造の上で代表的な点を5～8点選び、スプライン曲線で近似し長さを計測した。計測結果のヒストグラムを図3に示す。長さは平均 $20.4 \mu\text{m}$ で標準偏差 $2.9 \mu\text{m}$ であった。これは試料の λ DNA の塩基数 48,502 bp から予想²⁾される長さ $16.4 \mu\text{m}$ に比べると長い。この結果から単位長さ当たりの平均塩基数を求めると $2380\text{bp}/\mu\text{m}$ となる。

図4に線状構造の一つについて線方向に蛍光強度をプロットした結果を示す。線方向に強度分布を調べると40%程度の強度分布があることが確認される。

この傾向は他の線状構造についても同様であった。これは単位長さ当たりに固定された DNA の塩基数が異なるために生じた蛍光強度分布と推測される。単位長さ当たりの塩基数が異なる原因は、DNA を伸張固定する際に

水流によって伸張させているが、DNA が固定される場所や固定されるタイミングによって水流が変化すること、水流を受けるグロービュール状の DNA が基板上で上下することで水流から受ける力が変化すること、DNA の局所的な固さに分布があり伸張条件が異なること、などが考えられる。単位長さ当たりに固定された DNA の塩基数が局所的に変動すると仮定すると、先に述べた DNA 長の計測結果のばらつきも説明される。

GC コンテントの一次元分布を見るために、7-AAD での観察を行った。波長 532 nm のレーザーを照射して蛍光をイメージングした。図2(a) と同一試料の同一場所について観察した結果を図2(b) に示す。7-AAD では蛍光分子が若干ガラス基板上に吸着して蛍光発光するため DNA が存在していない部分でも若干の蛍光が観測されバックグラウンドが高くなる傾向がある。それに

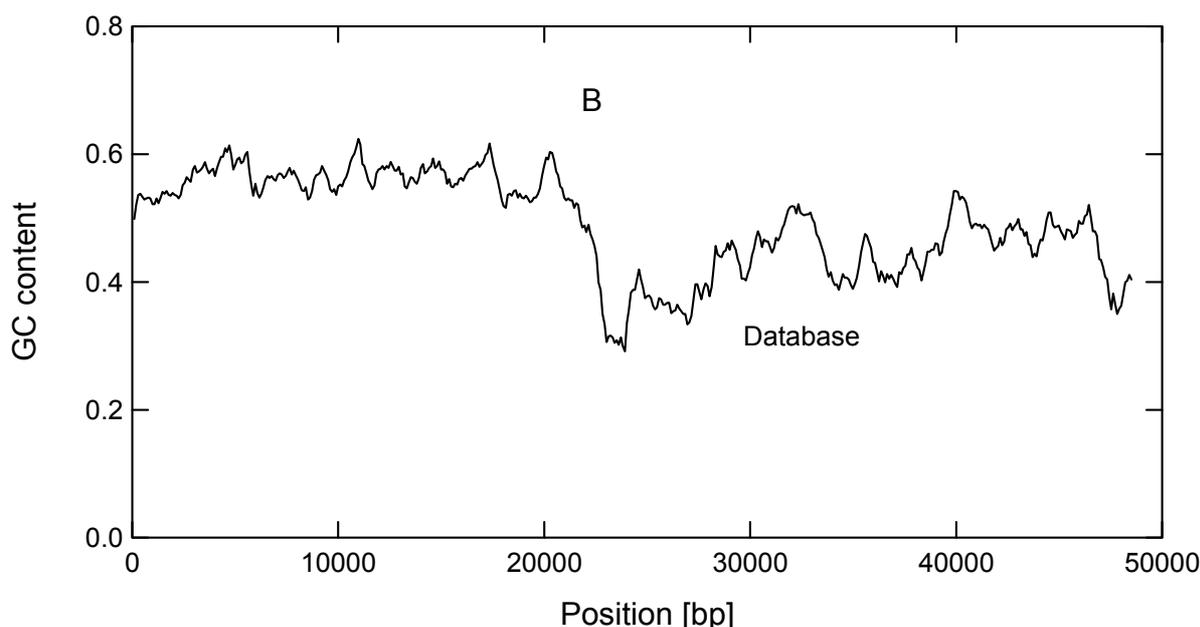


図7 λ DNA の塩基配列から求めた GC コンテント分布

対して YOYO-1 は DNA と結合したときに蛍光発光効率が高くなるので、バックグラウンドは比較的小さい。図 2 (a), (b) の中央部分に A-A' で示す DNA について延長方向の強度分布をプロットした結果を図 5 に示す。λ DNA では 5' 側に GC コンテントの高い領域があり中央付近で GC コンテントが低くなるのがデータベースから予想され、図 5 の結果はそれを反映した強度分布になっている。

4. GC コンテントの一次元分布の計測結果の解析

図 5 に示した線方向の計測結果を用いて解析を行った。GC コンテントは塩基配列情報から予想されるので、以下の方法で GC コンテントの一次元分布を求めて解析を行った。

塩基配列データ $U(k)$ から局所的な GC コンテントの分布を求めるには次の演算を行う。i 番目の塩基を中心にして、そこから 5' 末端方向へ $L/2$ 塩基、3' 末端方向へ $L/2$ 塩基の中に出現するグアニンおよびシトシンの個数を求めることで局所的 GC コンテント $R_{GC}(i)$ を定義する。局所的 GC コンテント $R_{GC}(i)$ は、

$$R_{GC}(i) = \frac{1}{L+1} \sum_{j=-L/2}^{L/2} G(j) \{ \delta_C(i+j) + \delta_G(i+j) \} \quad (1)$$

と求めることができる。ここで、

$$\delta_C(k) = \begin{cases} 1: U(k) = "C" \\ 0: U(k) \neq "C" \end{cases} \quad (2)$$

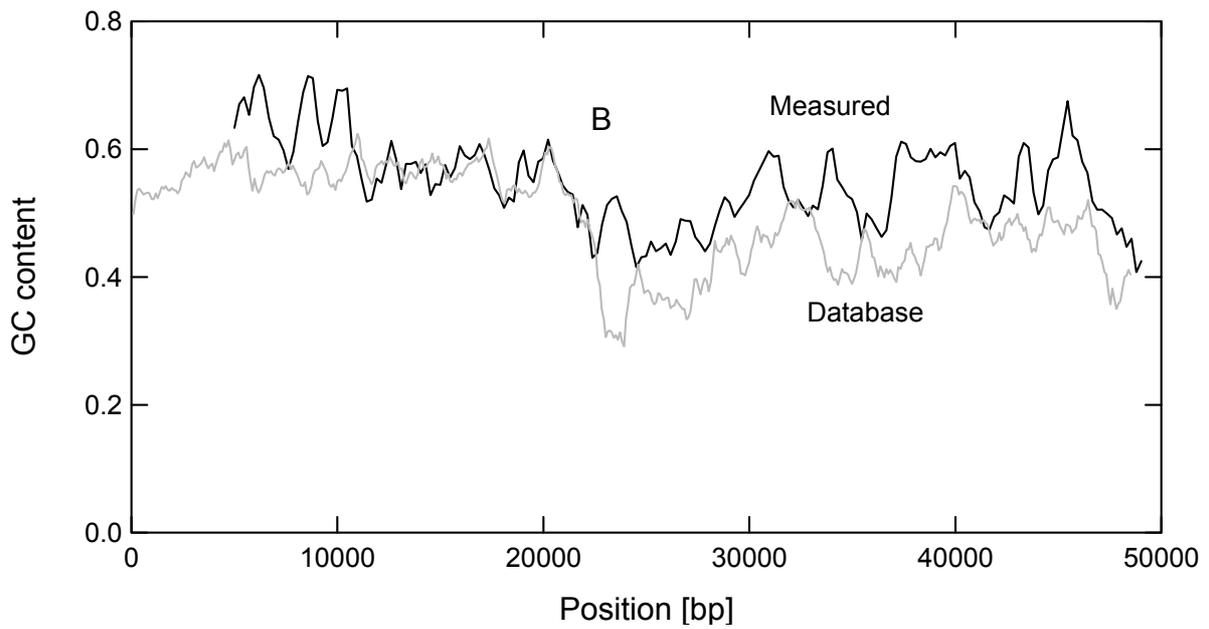


図7 長さの補正を行わなかった場合の解析結果

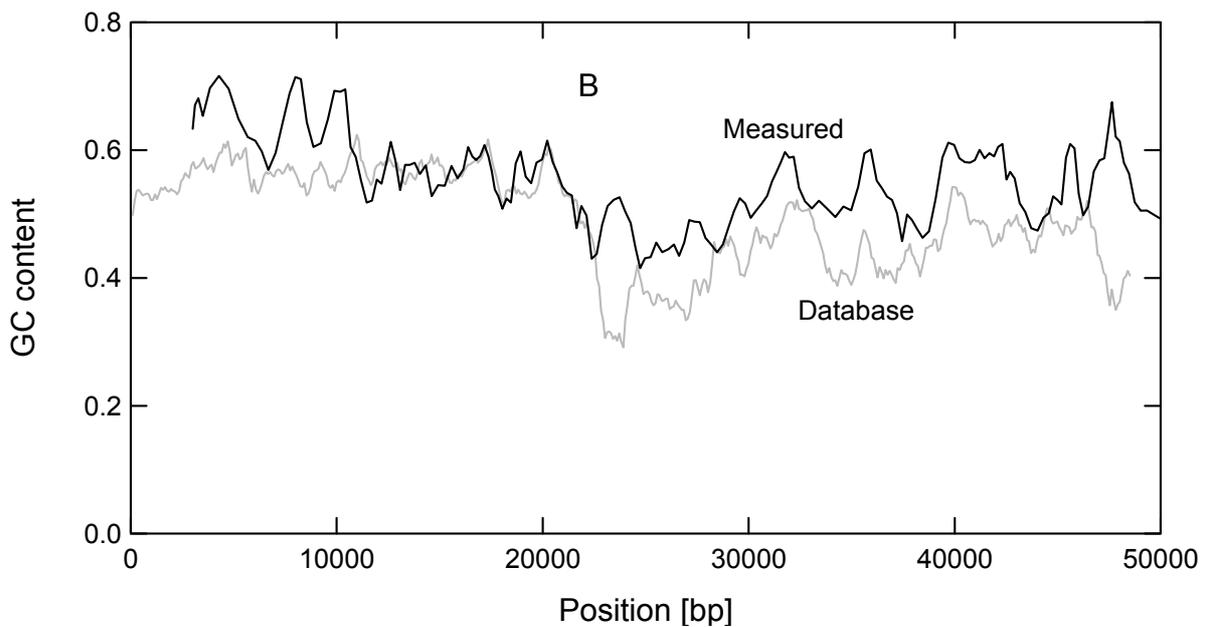


図8 長さの局所補正を行った場合の解析結果

$$\delta_G(k) = \begin{cases} 1: U(k) = "G" \\ 0: U(k) \neq "G" \end{cases} \quad (3)$$

である。U(k)はk番目の塩基を表す。Lを平滑化窓長さ、G(j)は窓関数で、顕微鏡光学系の点像広がりより決まる。実際の顕微鏡画像に対応させるには、塩基でR_{GC}(i)を長さの次元で表す必要がある。そこで、

$$R'_{GC}(x) = R_{GC}(f(x)) \quad (4)$$

と表す。 $f(x)$ は実座標と塩基数との関係を示す関数である。

λ DNA の塩基配列から求めた局所的な GC コンテント $R_{GC}(i)$ をプロットした結果を図 6 に示す。GC コンテントは中央付近を境に 5'側では高く、3'側では比較的低くなり、その境では 0.3 程度まで低くなっている。平滑化窓長さ L としては、先に求めた単位長さ当たり平均塩基数に対して顕微鏡の分解能である 250 nm を考慮して 600bp とした。

計測結果と比較するために $f(x)$ は $d x$ として計測結果と重ね合わせたものを図 7 に示す。データベースから予測された中央付近の GC コンテントの落ち込み B が一致するように重ね合わせてある。また各所で似た形状を持つことがわかる。特に落ち込み部分 B より左でよく一致している。しかしながら B よりも右側部分では類似なピークが連続した構造を持つが位置は一致しない。

そこでガラス基板上に固定された DNA が単位長さ当たりの塩基数が異なることを仮定して解析モデルを組み立てて解析を行った。ガラス基板上の DNA の単位長さ当たり塩基数は図 2 (a) に示す YOYO-1 による蛍光画像から推測される。蛍光強度 $I_{YOYO1}(x)$ と局所的な単位長さ当たり塩基数が線形な関係があると仮定すると、場所と塩基数の関係を示す関数 $f(x)$ は

$$f(x) = n_{\max} \int_0^x I_{YOYO1}(s) ds / p \quad (5)$$

と表される。ここで p は、

$$p = \int_0^{x_{\max}} I_{YOYO1}(s) ds \quad (6)$$

である。 n_{\max} 、 x_{\max} はそれぞれ総塩基数、画像上での DNA の延長である。また単位長さ当たりの塩基数が異なるとの仮定に基づき 7-AAD 強度分布 $I_{AAD}(x)$ から GC コンテント分布に対応する物理量を表す値 $I_{GC}(x)$ を求める式は、

$$I_{GC}(x) = \frac{I_{AAD}(x)}{I_{YOYO1}(x)} \quad (7)$$

で与えられる。実際の問題では $I_{GC}(x)$ は値を較正する必要がある。ここではデータベースから求められた R_{GC} の平均値 $\langle R_{GC} \rangle$ と $I_{GC}(x)$ の平均値 $\langle I_{GC}(x) \rangle$ が一致すると仮定して較正した。(7) 式を用いて解析した結果を図 8 に示す。先の解析結果と比べて B の右側部分のピークがシフトしてどのそれぞれのピークの一致具合がよく判別できる程度になっている。 I_{YOYO1} の強度に対する長さ補正のパラメータを適正化することでさらに一致度を高めることができると期待する。

5. おわりに

一分子 DNA の GC コンテントの一次元分布を計測手法について報告した。現時点では、一致度は必ずしも高くないが、GC コンテントの分布は計測できているとみなされる。一分子 DNA を伸張固定する際の長さ当たりの塩基数の違いを較正する方法についてさらに検討していく。

参考文献

- 1) T. T. Perkins, D. E. Smith, R. G. Larson and S. Chu, "Stretching of a single tethered polymer in a uniform flow," *Science*, Vol. 268, pp. 83-86 (1995).
- 2) M. D. Wang, H. Yin, R. Landick, J. Gelles and S. M. Block, "Stretching DNA with optical tweezers," *Biophys. J.* Vol. 72, pp. 1335-1346 (1997).
- 3) T. Sugiura, S. Kawata, and T. Okada, "Fluorescence imaging with a laser trapping scanning near-field optical microscope," *J. Microsc.*, Vol. 194, Pt. 2/3, May/June, pp. 291-294 (1999).
- 4) D. Axelrod, "Total internal reflection fluorescence microscopy in cell biology," *Traffic*, Vol. 2, pp. 764-774 (2001).