

# 一分子DNAのGCコンテンツ分布計測における長さ補正

## The length calibration on GC content distribution measurement for single molecular analysis of DNA

藤本 知之, 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

Tomoyuki FUJIMOTO, Tadao SUGIURA, Takeshi NAGAYA, Tetsuo SATO, Kotaro MINATO  
Department of Bioinformatics and Genomics,  
Nara Institute of Science and Technology

これまでに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コンテンツの一次元分布を計測できた。しかしガラス基板上に固定されたDNAの単位長さ当たりの塩基数が異なるため、データベースの予測と必ずしも一致しない。そこで、塩基数の違いを校正する新たな解析モデルとその適用結果について報告する。

We have analyzed single molecular DNA with single molecular imaging by evanescent field microscope and identified the DNA by comparing measured results of one dimensional distribution of local GC content to estimation from genome database. The number of base pair within the unit length of the DNA fixed on the glass substrates are different from estimation, however, measured results have not matched with genome database well. We report on the new analytic-model and the new application result considering calibration of the difference of the base pair number.

### 1. はじめに

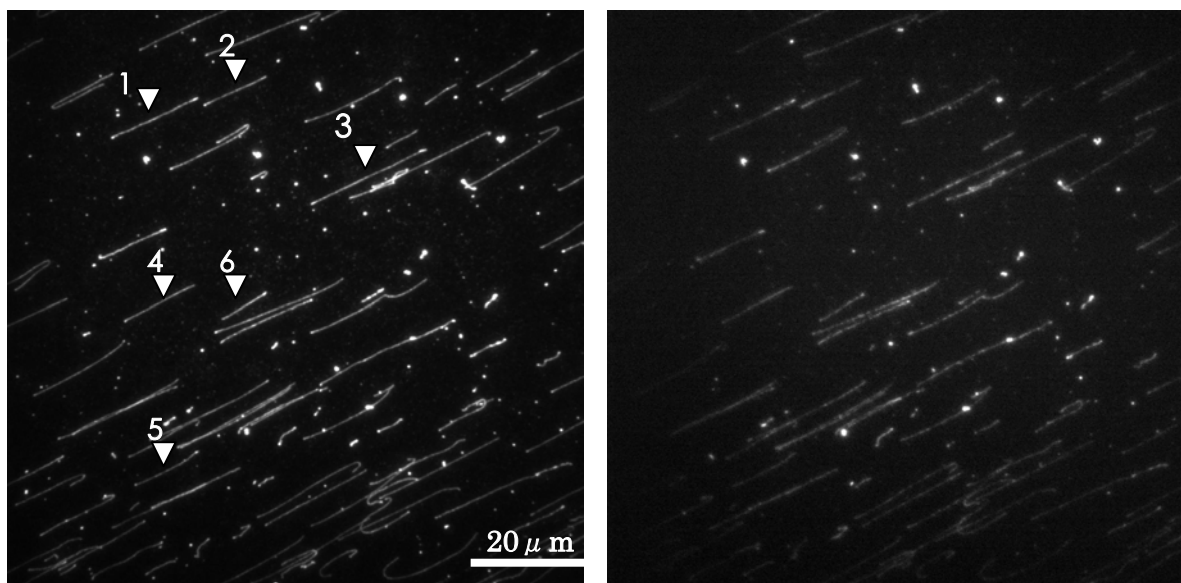
現在、エバネッセント顕微鏡等の光学技術の進展に伴い、一分子を対象とした高感度な計測が行われている<sup>1)</sup>。光を用いた計測は非破壊・非接触に対象物を観察できるといった特長を持ち、疾病の診断・予防・治療法の開発のためのゲノム解析研究に応用されている。近年加速するこのゲノム解析研究では、DNAの塩基配列情報を高速に同定することが求められている。塩基配列に関する情報を高速に得ることができれば、ゲノム配列データベースと比較を行うことによりウィルスの特定や亜種の発見が可能となり、早期の治療につながると期待できる。

そこで我々はDNAの局所的なGCコンテンツ分布は一樣ではない点に着目し、エバネッセント顕微鏡でDNAを一分子イメージングすることで、塩基配列情報の取得を試みた<sup>2)</sup>。ガラス基板上にDNAを伸張固定し、GC特異的な蛍光染色とDNA一樣な蛍光染色を組み合わせ、エバネッセント顕微鏡で一分子イメージングし、GCコンテンツの一次元分布情報を計測した。ガラス基板上へのDNA伸張固定方法は、スperlミジン水溶液の濃度によりDNAがグロビュール状態から伸張状態へ変化する性質を用い、ピペッティングによる水流を利用した<sup>3)</sup>。しかし、この方法で

は伸張時に DNA が水流から受ける力の変化, AT・GC 結合による結合力の違いに伴う伸張条件の変化により, 伸張固定した DNA の単位長さ当たりの塩基数に偏りが生じ, ゲノムデータベース上の予測と必ずしも一致しない.

データベース上の予測値と計測値の一致度を向上させるためには, 単位長さ当たりの塩基数の違いを較正する必要がある. そこで DNA を一様に蛍光染色して得られる画像の画素値情報を用いることで, 単位長さあたりの塩基数の違いを較正する手法を考案した.

本稿では GC コンテントの計測手法について述べた後, 単位長さあたりの DNA の塩基数の較正法及び計測結果を示し, 従来法との比較について報告する.



(a) YOYO-1 蛍光画像

(b) 7-AAD 蛍光画像

図 1. 伸張固定した DNA の観測結果

## 2. GC コンテント計測手法

48,502bp の  $\lambda$  DNA を用い, 蛍光分子 7-AAD により GC 特異的に蛍光染色した後, YOYO-1 により DNA 一様に蛍光染色する<sup>4)</sup>. この 2 重染色した DNA をガラス基板上に伸張させ, 各蛍光分子をエバネッセント顕微鏡により励起させ撮影を行う. これら蛍光分子の励起波長  $\lambda_{ex}$  と蛍光波長  $\lambda_{em}$  は, 7-AAD では  $\lambda_{ex}=550$  nm,  $\lambda_{em}=650$  nm, YOYO-1 では  $\lambda_{ex}=490$  nm,  $\lambda_{em}=510$  nm であり, 波長の違いから容易に識別可能である. エバネッセント顕微鏡は, エバネッセント状態となった光を励起光として利用する蛍光顕微鏡で, バックグラウンドの影響を受けずに基板表面近傍のみを観察することができる<sup>5)</sup>.

図 1(a)に YOYO-1 を励起させた画像, (b)に 7-AAD を励起させエバネッセント顕微鏡により撮影した画像を示す. 各画像は 12bit 濃淡画像で, 1 辺は  $100 \mu\text{m}$ , 1024 画素となっている. 7-AAD の蛍光画像はコントラストが低く DNA の判別能が劣るため, DNA の抽出には YOYO-1 蛍光画像を用いる. YOYO-1 蛍光画像を用い DNA 線分の尾根線を抽出し画素値  $f_{YOYO-1}$  を求める. 尾根線の抽出は, DNA の両端 2 点を結ぶ直線に垂直な複数の線上で最も高い画素値を求めることで行

う．隣り合う垂直な線の間隔は画像上で求めたユークリッド距離の単位長さとしている．続いて 7-AAD 蛍光画像上で  $f_{YOYO-1}$  と同じ位置の画素値  $f_{7-AAD}$  を得る．そして，(1)式により GC コンテントを算出する<sup>1)</sup>．この操作を選択した各 DNA のピクセル毎に行っていくことで，1 本の DNA の GC コンテントを求めることができる．なお DNA 選択時における尾根線抽出では GC コンテントを求めるための間隔は一定とは限らないが，本報告においてはこのサンプリング間隔を一定と扱う．

$$\text{GCcontent} = \frac{f_{7\text{-AAD}}}{f_{\text{YOYO-1}}} \quad [\%] \quad (1)$$

また 7-AAD と YOYO-1 の蛍光強度の違いについては， $\lambda$  DNA の GC コンテントの理論値と計測値の平均値が一致すると仮定し補正している．

### 3. 単位あたり長さの較正法

図 2 に図 1(a)の 1 より得た DNA の画素値分布，図 3 に(b)より得た同じ位置にある DNA の画素値分布を示す．蛍光分子 YOYO-1 は DNA を一様に染色するため，図 2 の画素値分布も一様となっているべきである．しかし DNA の伸張度合いが一定でないことから，画素値分布に偏りが生じている．そこで，この画素値分布を基に DNA の伸縮を推定する．また図では DNA 両端の画素値が他の場所と比べ極端に高くなっている．この理由としては，DNA がグロビュール状態のまま伸張せずにガラス基板上に固定されているものと考えられる．そのため伸張を推定する前処理として両端を対象領域外とし，中央部分の情報のみを用いて単位あたり長さの較正を行う．

YOYO-1 の画素値が高い場所は 1 画素に含まれる蛍光分子の数が多く，すなわち含まれている塩基数も多いと考えられ収縮しているものとする．逆に画素値分布の低い場所は蛍光分子が少ないため DNA 分子数が少なく伸張していると考えられる．ここでは伸張を実現するために以下の 3 種類の伸張法について比較検討を行った．

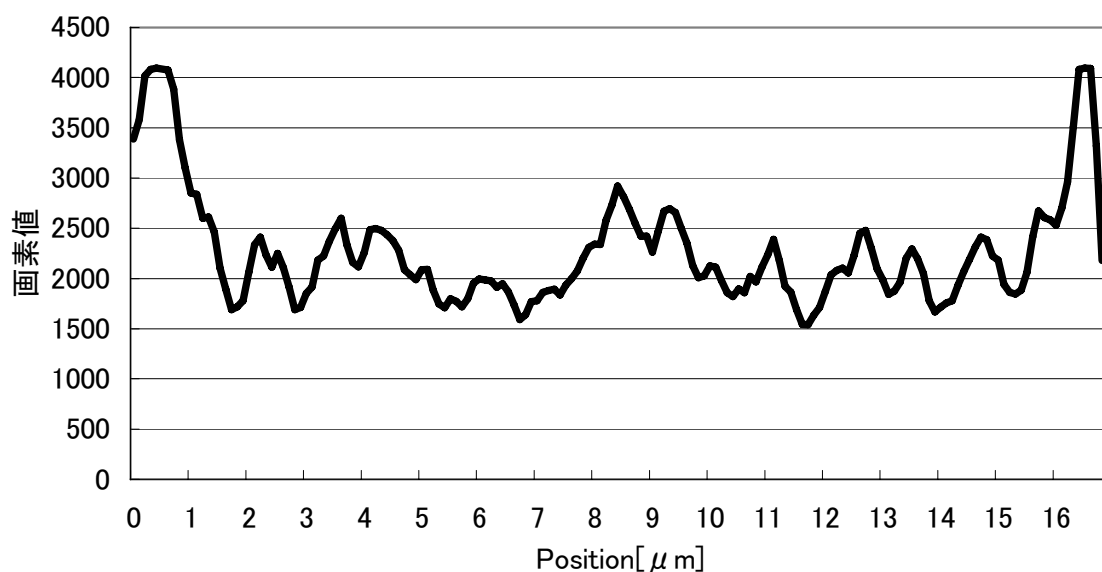


図 2. YOYO-1 の画素値分布

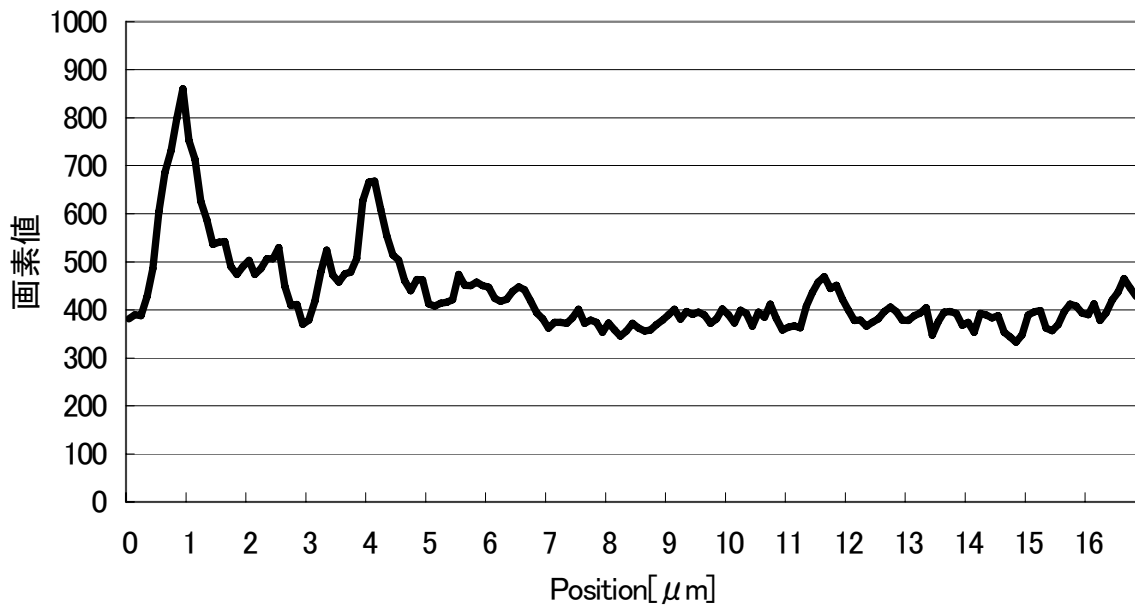


図 3. 7-AAD の画素値分布

1). YOYO-1 の画素値に対し線形に伸張

YOYO-1 の画素値の最大値(max)と最小値(min)を求め、伸張度合いを画素値に対し線形に変化させる。DNA が伸張しているとみなせる画素値の低い位置は単位長とし、画素値が最大値に近づくにつれ長さを  $n$  段階で線形に変化させる。ある画素値( $f_{YOYO-1}$ )を持つ点の伸張度(ex)は(2)式で計算する。今回は  $n=10$  とし計算を行った。

$$ex = \frac{f_{YOYO-1} - \min}{\max - \min} \times n \quad (2)$$

このとき、7-AAD の画素値分布に対しても YOYO-1 の画素値分布を基に同様の伸張操作を行う。また伸張後の 7-AAD の画素値( $f'_{7-AAD}$ )は(3)式のように伸張度が増すごと 10%ずつ値を低くするように操作を行っている。画素値が最大値なら長さは 10 倍、 $f'_{7-AAD}$  は  $f_{7-AAD}$  の 10%の値となる。

$$f'_{7-AAD} = f_{7-AAD} - f_{7-AAD} \times 0.1(ex - 1) \quad (3)$$

2). YOYO-1 の画素値に対し指数的に伸張

上述した手法 1 と同様に YOYO-1 の最大値と最小値を使用し、(4)式のように伸張度を指数的に変化させた。

$$ex = ab^n \quad (4)$$

ここで  $a = \exp \frac{-\min \times \ln(n)}{\max - \min}$  ,  $b = \frac{\ln(n)}{\max - \min}$  で求めた指数曲線のパラメータとする。

3).YOYO-1 の単位画素値毎に 1 ずつ伸張

YOYO-1 の画素値が単位画素値を超える毎に 1 ずつ伸張させる。単位画素値を 500 とし、画素値が 500 以下では単位長、以降画素値が 500 増える毎に単位長の 2 倍、3 倍とし、最大値である 4095 では 9 倍となり、 $n=9$  となる。この方法は上記 2 つの最大値と最小値を用い、相対

的に伸張させる手法とは異なり，YOYO-1 の画素値分布が異なっても画素値が同じであれば伸張度合いは一定となる。

上記 3 手法を用い DNA の単位長さの違いを再計算することで伸張させた後，今回使用した  $\lambda$  DNA の理論上の長さである 約 16[ $\mu$ m]まで全長 DNA を線形に縮めることで単位あたり長さの較正を行っている。図 4 は  $\lambda$  DNA の GC コンテンツの理論値で，エバネッセント顕微鏡の分解能を考慮し単位長さあたりの塩基数を 600bp と仮定している<sup>2)</sup>。 $\lambda$  DNA の GC コンテンツは図中○のように中央付近に特徴的な落ち込みが見られる。この特徴的な落ち込み部分を一致させた後に全長 DNA を縮めている。

#### 4. 計測結果

撮影した DNA の 6 本について補正なし，及び伸張を考慮したものについて比較を行った。6 本の DNA のうち最短は約 11  $\mu$ m，最長で約 25  $\mu$ m，平均は約 16  $\mu$ m となった。

理論値と計測値の類似度を判断するために(5)式の相関係数を求めた。

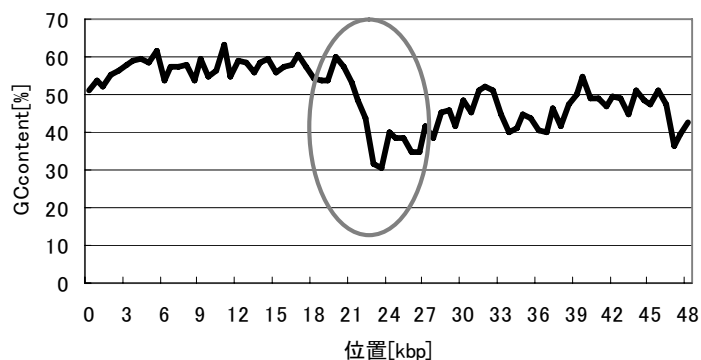


図 4.  $\lambda$  DNA の GC コンテンツ

$$C = \frac{\sum_{i=1}^N \{ (DB_{GC}(i) - \overline{DB}_{GC}) (M_{GC}(i) - \overline{M}_{GC}) \}}{\sqrt{\sum_{i=1}^N \{ (DB_{GC}(i) - \overline{DB}_{GC})^2 \}} \sqrt{\sum_{i=1}^N \{ (M_{GC}(i) - \overline{M}_{GC})^2 \}}} \quad (5)$$

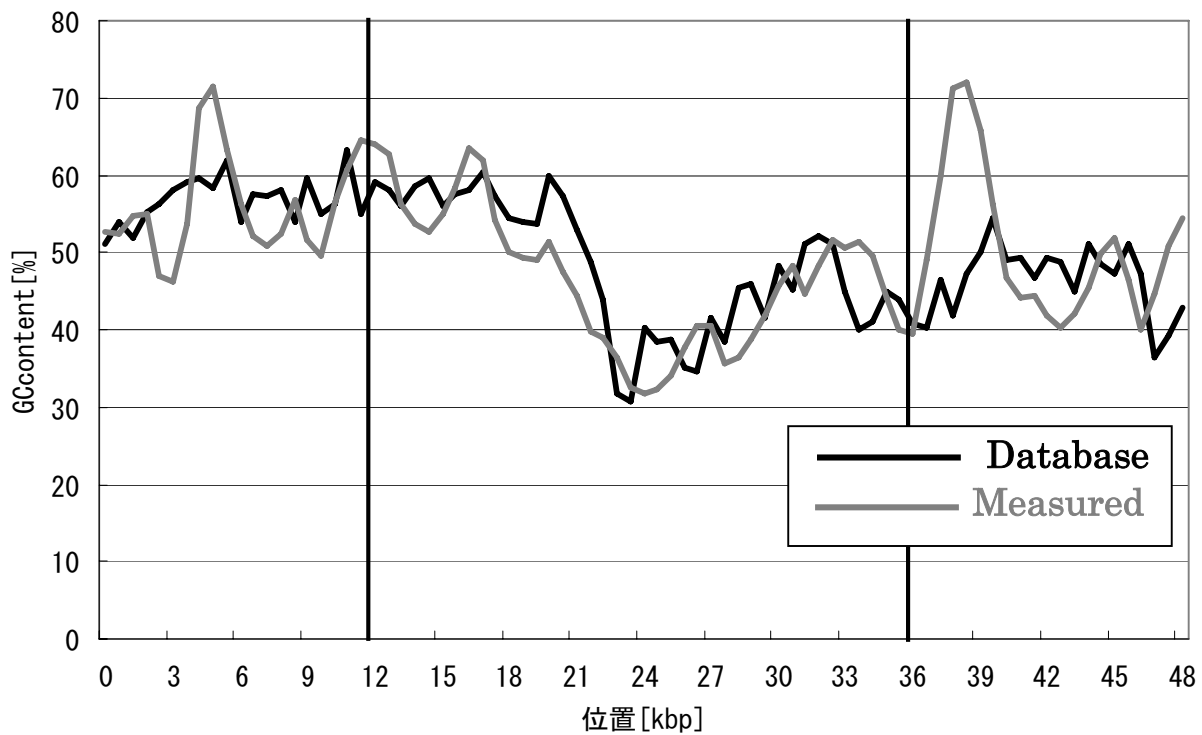
ここで， $DB_{GC}(i)$ は  $i$  番目の理論値の GC コンテンツ， $\overline{DB}_{GC}$ は理論値の GC コンテンツの平均値， $M_{GC}(i)$ は  $i$  番目の計測値の GC コンテンツ， $\overline{M}_{GC}$ はその平均値となる。また  $N$  は理論値のデータ数となる。(5)式を用いて図 1(a)の DNA，1 から 6 番の比較を行った結果を表 1 に示す。また，DNA1 の GC コンテンツを図 5 に，DNA6 の GC コンテンツを図 6 に，(a)は較正なし，(b)は手法 1 を適用したものを示す。

表 1. GC コンテンツの理論値と計測値の比較

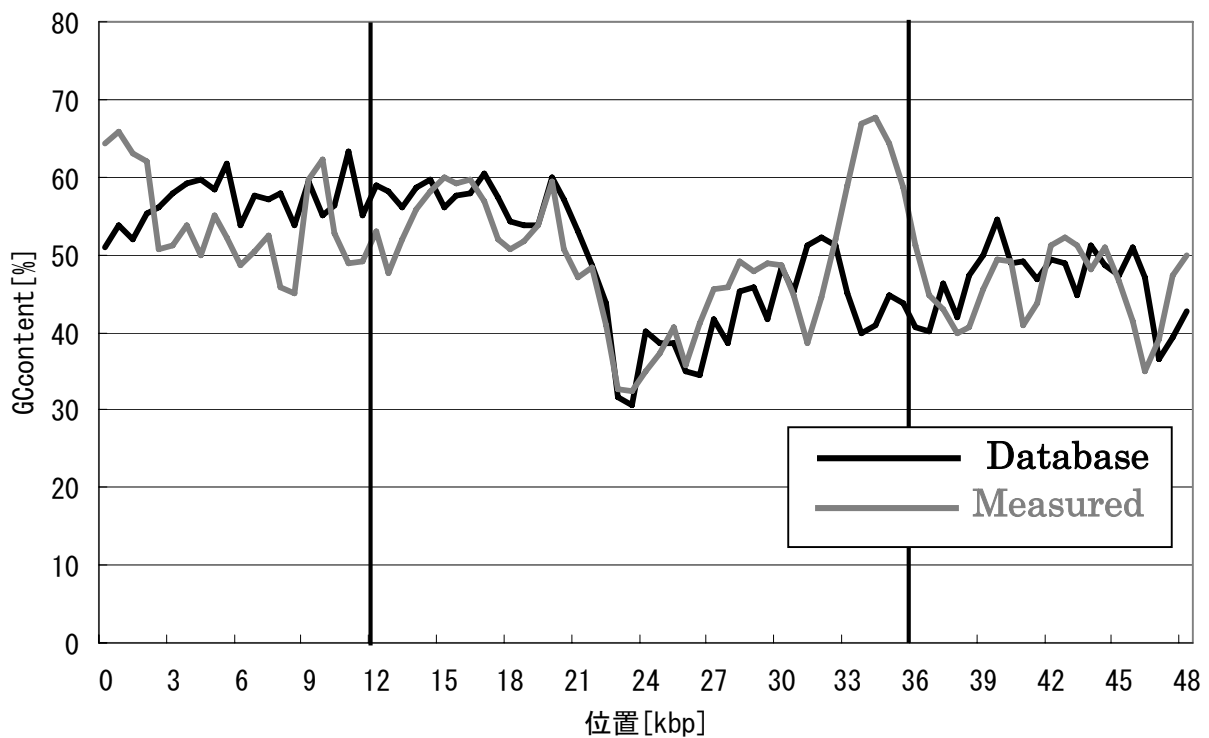
DNA	相関係数			
	手法1	手法2	手法3	較正なし
1	0.483	0.595	0.582	0.619
2	0.711	0.699	0.736	0.623
3	0.320	0.098	0.231	-0.052
4	0.174	-0.280	0.028	0.050
5	0.494	0.383	0.444	-0.167
6	0.523	0.549	0.573	0.260

表 2. 中央部のみの比較

DNA	相関係数			
	手法1	手法2	手法3	較正なし
1	0.522	0.683	0.658	0.810
2	0.752	0.703	0.794	0.650
3	0.172	-0.136	0.008	0.135
4	0.458	0.113	0.342	-0.276
5	0.562	0.132	0.181	0.257
6	0.540	0.518	0.477	-0.101

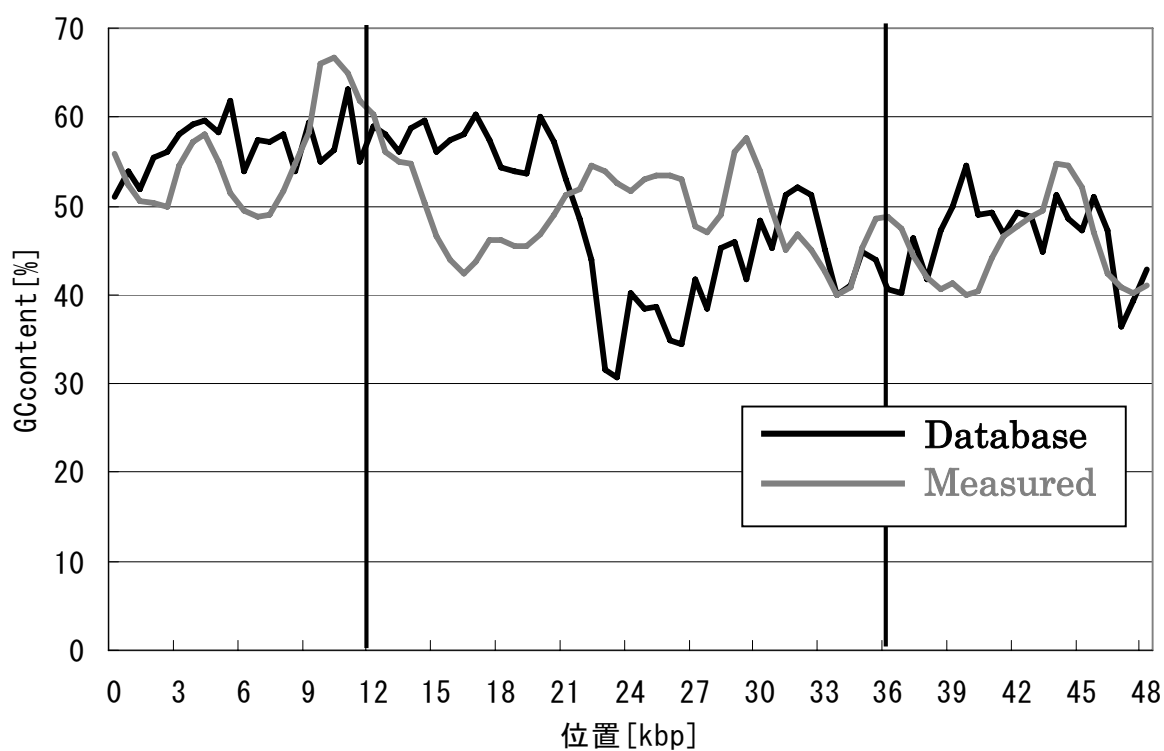


(a) 理論値と校正なしデータの比較

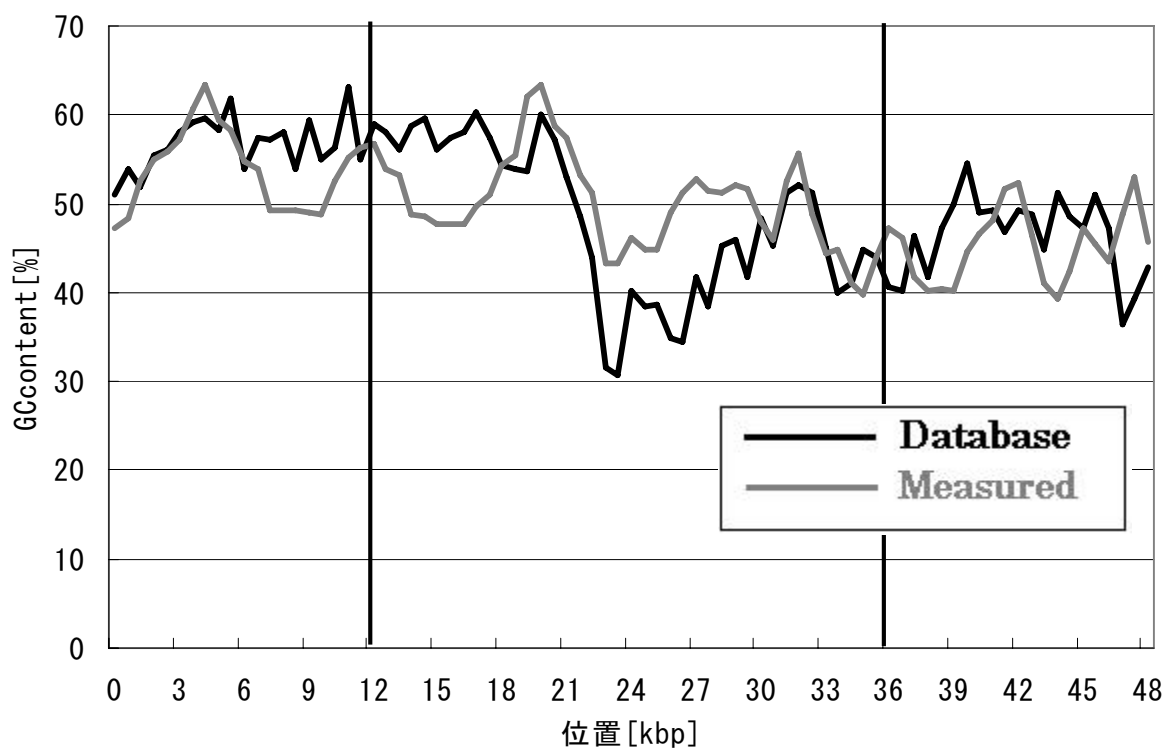


(b) 理論値と手法 1 の比較

図 5. 図 1(a)中 1 の GC コンテンツの理論値と計測値の比較



(a) 理論値と較正なしデータの比較



(b) 理論値と手法1の比較

図6. 図1(a)中6のGCコンテンツの理論値と計測値の比較

表 1 より、伸張を考慮した場合に係数の向上が見られたのは 6 本中 5 本となった。図 1(a)中 1 に限り較正しない場合の係数が高かった。また、図 1(a)中 6 は伸張を考慮することで相関係数が向上している。しかし、全体的に係数は高いとはいえない。原因として、中央部の落ち込みのある領域は一致しているが、両端での細かな差異が考えられる。そこで、中央部 12kbp から 36kbp 付近に着目し、再計算した相関係数を表 2 に示す。

表 2 と表 1 を比べると半数以上の係数で値の向上が確認できた。特に図 1(a)中の 1 と 2 については全てにおいて向上が見られた。しかし図 1(a)中 3 については較正した値全てが表 1 より減少している。また、手法ごとに見ると 1 では表 1 に比べ全体的に係数が向上しているが手法 2 及び 3 では共に半数の向上にとどまった。ここでは、指数的に伸張させる手法 2 を適用した結果よりも線形に伸張させる手法である 1, 3 を適用する方が高い係数を示した。

## 5. まとめ

DNA 伸張固定時に生じる単位長さあたりの塩基数の違いを較正するために、一分子 DNA の GC コンテント計測時における伸張を考慮し、単位あたり長さを較正した結果について述べた。伸張を考慮し単位あたり長さを較正することでデータベースとの類似度を向上させることができた。また、伸張方法については、指数的に伸張させるよりも線形的に伸張させるほうがより類似度が向上した。本報告では伸張させたデータを収縮させる際に線形手法を用いたが、実際の DNA の伸張は非線形であると思われるため、GC コンテントを基に各分子の結合力の違いを考慮した伸張法などを検討する必要があると思われる<sup>6)</sup>。また、エバネッセント顕微鏡による照明場が一樣ではない点や、一定と仮定している尾根線のサンプリング間隔の補正を行うことで、さらなる GC コンテントの類似度の向上が期待できる。

## 参考文献

- 1) Y.Harada, T.Funatsu, K.Murakami, Y.Nonoyama, A.Ishihara, and T.Yanagida, "Single-molecule imaging of RNA polymerase-DNA interactions in real time", *Biophysical Journal* Vol.76 February 1999, pp709-715 (1999).
- 2) 杉浦忠男, 長屋岳志, 佐藤哲大, 湊小太郎, "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", 情報処理学会研究報告, 2006-BIO-004, Vol. 2006, No. 13, pp25-32 (2006).
- 3) Smith, S., Y.Cui, and C.Bustamante, "Direct mechanical measurements of the elasticity of single DNA molecules by using magnetic beads", *Science*, Vol. 258, pp1122-1126 (1992).
- 4) R.Zucker, A.Keshaviah, O.Price, and J.Goldman, "Confocal laser scanning microscopy of Rat Follicle development", *The Journal of Histochemistry & Cytochemistry* Vol. 48(6), pp781-791 (2000).
- 5) D.Axellrod, "Total internal reflection fluorescence microscopy in cell biology", *Traffic*, Vol.2, pp. 764-774 (2001).
- 6) C.Bouchiat, M.D.Wang, J.-F.Allemand, T.Strick, S.M.Block, and V.roquette, "Estimating the persistence length of a worm-like chain molecule from force-extension measurements", *Biophysical Journal* Vol.76, January 1999, pp409-413 (1999).