



# 大腸菌における $\alpha$ 配列の過剰頻度と 4 コドン頻度の関連に関する解析

政策メディア研究科 鷓野 れいな

## Abstract

大腸菌において、 $\alpha$  配列 (5'-GCTGGTGG-3') は RecBCD により認識され相同組換えを促進するホットスポットである。我々は以前、 $\alpha$  配列の方向が複製側に偏る傾向は、リーディング鎖の G 塩基偏りによって生じており、その役割のために生じたのではないという結果を報告した。そこで  $\alpha$  配列が現在のような配列になったのは進化の過程において、組換え酵素が都合の良い (=数の多い) 配列を選択したためであると考え、現在のような GCTGGTGG という配列が大腸菌ゲノムにおいて  $\alpha$  配列として選択されたのは多用されるコドンを使用したという仮説を立てた。今回、我々は大腸菌ではロイシン・バリンというアミノ酸が多用されていることと、 $\alpha$  配列の約 70% が ctg-gtg のロイシン・バリンの読み枠で読まれるという二つの解析結果から、ゲノム上に高頻度で存在するロイシン [ctg] + バリン [gtg] という 6 塩基配列の両側に g を配するようなコドンが高頻度で現れて大腸菌  $\alpha$  配列を形成していると推測し、4 コドン頻度のマルコフ解析を行った。その結果、 $\alpha$  配列自体は配列を増加させるような選択圧を受けてはおらず、コドン (アミノ酸) の影響によるものであるという結果が得られた。これにより、大腸菌の  $\alpha$  配列頻度が高次のコドン頻度で説明できることがわかった。

# 1 はじめに

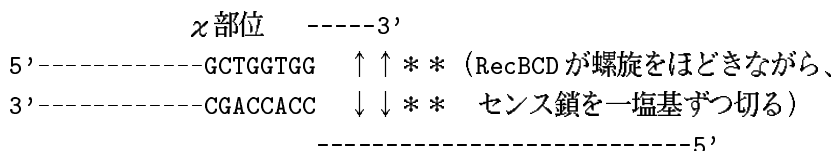
## 1.1 $\alpha$ (カイ) 配列とは?

大腸菌において  $\alpha$  配列とは 8bp の配列で、次のような配列を指す名称である。



$\alpha$  配列は、一塩基ずつ切断していく酵素「エキソヌクレアーゼ V」としての役割を果たすタンパク「RecBCD」の標的部位である。

エキソヌクレアーゼ V が  $\alpha$  配列の下流の DSB(double strand break) に結合すると DNA を 3' から 5' の方向に一塩基ずつ切りながら  $\alpha$  配列に向かって移動し、 $\alpha$  配列に達すると、その配列の 4~6 塩基右側でヘリカーゼの機能に変化する。 $\alpha$  配列の役割は遊離の末端を作り出すことにあり、この遊離末端がゲノムの相同な領域と相同組換え (homologous recombination) を起こすと言われている [?].

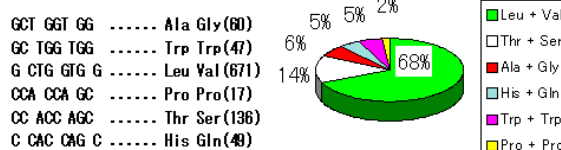


この 8 塩基の  $\alpha$  配列は、大腸菌 K-12 株には 1009 個存在している [3][?]. これは 8 塩基配列としてはゲノム中で高頻度に出現する配列である。この  $\alpha$  配列は種によって配列が異なっていることが実験により証明されている [4][5][6][7][8]. またこれらの  $\alpha$  配列は、ORF 内に比較的高い割合で存在している。大腸菌の場合、 $\alpha$  配列は 8 塩基であるため、表配列・裏配列合わせて 6 通りの読み枠が考えられるが、このうち約 7 割が Leu-Val(ctg-gtg) の読み枠に偏っている。この理由として大腸菌における高頻度使用コドン (or アミノ酸) として [ctg][gtg] が用いられていることが考えられる。

### Escherichia coli Chi



◆  $\alpha$  配列には 6 通りの読み枠がある ◆



このように大腸菌  $\alpha$  配列の多くは Leu-Val に翻訳される。

## 1.2 これまでの仮説

$\alpha$  配列には方向性があり、大腸菌ゲノムの場合では 75% が複製方向と同一方向に向いている [3]. それらは複製に伴って DNA の損傷を修復するために組換えが生じているためという考えが一般的には考えられている [11][12]. これまでの我々の解析の結果で  $\alpha$  配列の複製方向への偏りはゲノム上の G・C の偏り (GC skew[3]) で説明できることが分かった [13]. またさらに大腸菌  $\alpha$  配列は (複製よりもむしろ) 転写方向に向いているという報告もあり [15][16]. 転写についても同様の解析を行なったところ GC skew,

AT skew によりこれらの相関が説明できた。 $x$ 配列が複製・転写方向を向くのは、ゲノム中のグアニンの遍在の結果 (GCskew) (転写方向に関してはアデニンも影響あり) によるものと結論づけた。つまり $x$ 配列が複製・転写方向を向くという選択圧がゲノム全体にかかっていたわけではないことが示唆される。

$x$ 配列が複製方向に向くような力がかかってきたわけではなく、組換え酵素 RecBCD (AddAB, RexAB) が条件にあう配列を選んで来たのではないかと考えられる。ある種において $x$ 配列を増やすなどの圧力が働いて、 $x$ 配列が出来てきたのではなく、組換え酵素が条件に合った配列を $x$ 配列として選択するように進化したのではないだろうか? ある配列の方向性を変えたり、数が増えるような圧力をゲノム全体にかけるとしても組換え酵素の認識配列を変える (つまりタンパク側を変化させる) 方が圧倒的に効率的だろう。組換えを促進する配列として最も重要な条件は、ゲノムのどこの部位でも組換えられるだけの $x$ 配列の数を持つことであろうと考えられる。

故に $x$ 配列の特徴や出現頻度を考慮して、各種において次のような仮説を立てた。

- 大腸菌：最初から数が多い配列を選んだ。[5'-gctggtgg-3']
- 枯草菌：認識配列を5塩基にした。[5'-agcgg-3']
- インフルエンザ菌：複数の配列を認識可能にした。[5'-gNtggtgg-3', 5'-g(c/g)tgagg-3']

### 仮説の問題点

大腸菌の $x$ 配列の出現頻度は予想されるものよりも遥かに高いことは周知の事実である [3][17]。我々のこの仮説は大腸菌における高出現頻度を十分に説明できていない。 $x$ 配列の機能によって高い出現頻度になった可能性が未だに捨てきれない。

これまで、塩基頻度のマルコフ解析を元にした予測値との比較では $x$ 配列の高出現を説明できなかった。さらに大腸菌で多用される [ctg] コドンを用いた1コドン頻度や2コドン頻度を元にした予測値との比較でさえも、大腸菌 $x$ 配列の高出現頻度を説明しきれいかなかった [17]。大腸菌 $x$ 配列は非常に高い割合で出現しているということもあり、進化的圧力によりこの $x$ 配列が高頻度に出現するような選択圧がかかってきたという議論が一般的になされている。

しかしながら我々はこの大腸菌において高頻度出現である $x$ 配列が「選択圧によるものではない」ということを証明するために、3コドン以上のマルコフ解析において大腸菌 $x$ 配列の高出現頻度の期待値と実測値の算出を行ない、 $x$ 配列は多用されるコドンを RecBCD タンパクがたまたま利用したに過ぎないことを数値的に証明する。

## 2 手法

$x$ 配列のコアとなる二つのアミノ酸を中心として、前後にどのような塩基が位置するかをマルコフ解析を用いて調べる。

gtggcgtgctggtggatgagtga



di-codon を中心にして前後にグアニンが来る確率を、一次マルコフにより求める。

それぞれ大腸菌  $\alpha$  配列のコアとなる 2 アミノ酸に対して、4 コドンマルコフ解析で期待値を求め、実測値と比較し O/E 値を算出する。その際、コアとなる 2 アミノ酸では、

- (1)  $\alpha$  配列と「同じ」コドンを使用。
- (2)  $\alpha$  配列「以外」のコドンを使用。のどちらであるかを分けてカウントする。

仮に  $\alpha$  配列が高頻度に出現するような進化的な圧力がかかっていたとするならば、同じ 2 アミノ酸でも  $\alpha$  配列と同じコドンの O/E 値が他の同義コドンよりも高くなっているはずである。

### 3 結果

L-Val 周囲の塩基と O/E 値

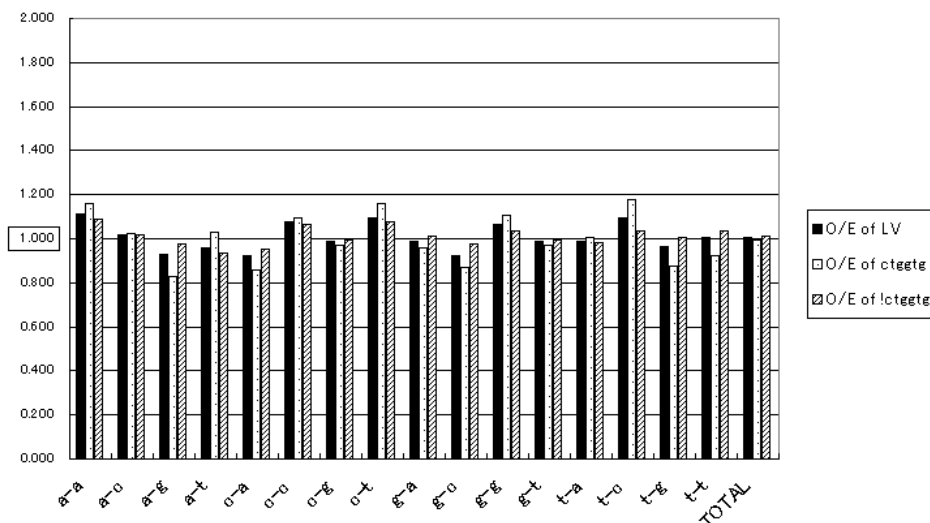
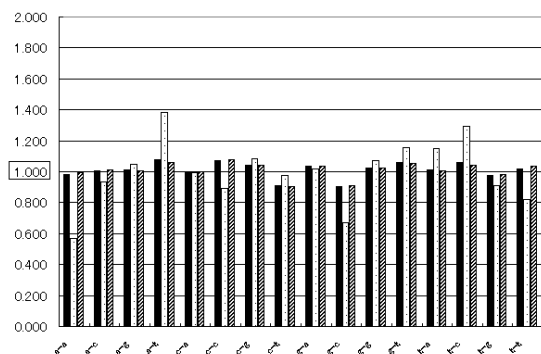


図 1 : Leu-Val の読み枠における前後の塩基の O/E 値

大腸菌  $\alpha$  配列の読み枠の中で最も多く 7 割弱を占める。前後にどの塩基が来てもほぼ一様に O/E 値は 1.0 付近を示している。また  $\alpha$  配列となっている [ctg]-[gtg] と、それ以外の同義コドンと場合分けして O/E 値を算出したが、どちらも大差がないことから、 $\alpha$  配列のみに選択圧がかかっていないことがわかる。

Ala-Gly(gct-ggt)周囲の塩基と O/E 値



Trp-Trp(tgg-tgg)周囲の塩基と O/E 値

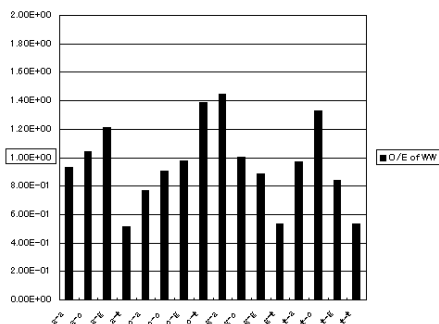


図 2・3 : Ala-Gly, Trp-Trp の読み枠における前後の塩基の O/E 値

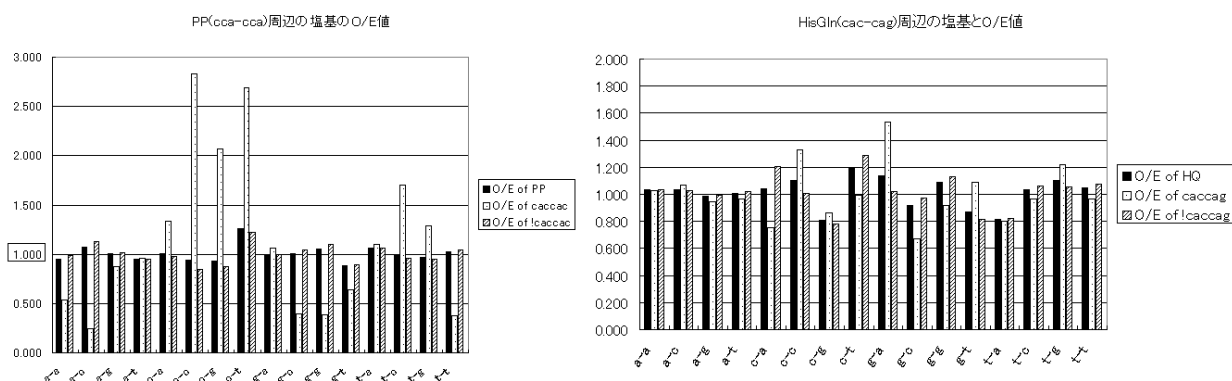


図 4・5 : Pro-Pro, HisGln の読み枠における前後の塩基の O/E 値

他の読み枠はそれほど割合は高くないが、コントロールのデータとして表示した。Trp-Trp に関しては他の読み枠が終止コドンとなるため  $x$  配列と同じ [tgg][tgg] のコドンのみとなっている。

Pro-Pro では多少のバイアスがかかっており、2 アミノ酸の前に C が来やすい傾向にあるようだ。

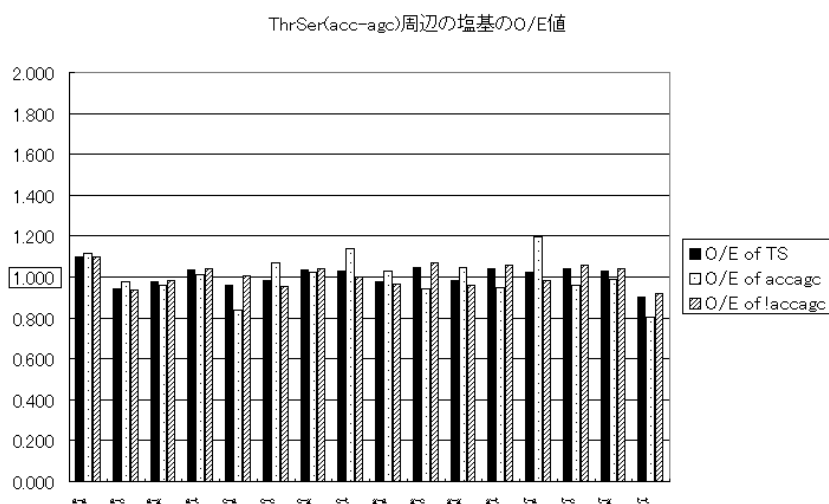


図 6 : Thr-Ser の読み枠における前後の塩基の O/E 値

$x$  配列がアミノ酸に翻訳去れる際に、Leu-Val の次に割合の高い (14%) Thr-Ser の読み枠では、やはり O/E 値はどれも 1.0 に近く偏った傾向は見られないことがわかる。

#### 4 結果その 2 : 組換えタンパク (RecBCD) の認識配列

組換え酵素としての RecBCD サブユニットで働くが、その中でも RecC タンパクは  $x$  配列を認識する部位として知られている [18] [19]。

- インフルエンザ菌 : 認識配列を甘くすることで数を増やした。

我々の仮説では上記のように考えたが、どこの部位が変異することで認識配列を甘くすることが可能となるのかを調べるため、大腸菌とインフルエンザ菌とでアミノ酸のアライメントをかけた。それぞれの RecC の相同性は 35% であった。これは 10 塩基ごとの Window 幅でどの領域が 2 種間でどれくらいの相同性を持っているかをグラフ化したものである。相同な領域にはいくつかのピークがありこの辺が RecC の立体構造に関連して保存されているのではないかと考えられる。この○で示したピークは非常に相同性があり、約 20 アミノ酸ほどの領域が保存されている。このピークの直前が  $\chi$  配列の配列認識部位ではないかと言われている。Arnold, Handa らの論文によると recC(1004) 株ではある領域の 9 アミノ酸が他のアミノ酸に置換しており  $\chi$  配列の認識配列通常の 8 塩基配列ではなく 11 塩基の  $\chi$  配列に類似した配列を組換え促進配列として認識したという報告がなされている [19]。

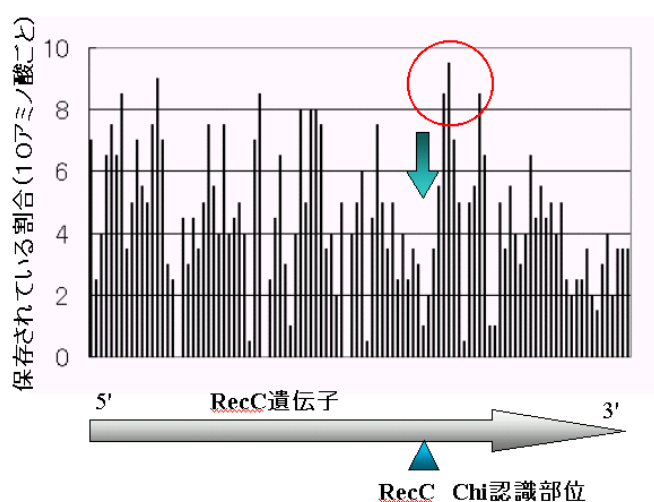


図 7: 大腸菌とインフルエンザ菌の RecC の保存度

予想通り 9 アミノ酸部位には変異がある。後ろの部分が (グラフ中○で記した部位) 比較的高度に保存されていることが分かる。

```
Ecoli: 629  GDAVPLSLRDELAQRLDQERISQRFLAGPVNICTLMPMRSIPFKVVCLLGMNDGVYPRQ 688
          + +  ++ D +  +L+    S +FLAG VN CTL+PMRS+PFKVVCLLGMND  YPR
Hinf : 637  NEELQAEVIADVITMQLEDAPNSLKFLAGKVNFCITLLPMRSVPFKVVCLLGMNDADYPRT 696
          (1) ~~~~~ (2) →ここより下流が比較的保存されている
```

```
Ecoli: 689  LAPLGFDLMSQPKPKXXXXXXXXXXYLFLLEALISAQKLYISYIGRSIQDNSERFPSVLV 748
          P  FDLM   +                YLFLLEAL++A+  YISY+GRSI DN  + PSVLV
Hinf : 697  QTPNSFDLMQYHYQKGDVRDRDDRYLFLLEALLAARDYCYISYVGRSITDNQPKEPSVLV 756
```

(1) あまり保存されていない 9 アミノ酸部位 (recC1004 の変異箇所)

-> これが、インフルエンザ菌と大腸菌の  $\chi$  配列の違いを意味している

(2) 比較的保存されていた (変異 9 アミノ酸部位の後ろ) 20 アミノ酸部位

-> インフルエンザ菌と大腸菌の  $\chi$  配列が似ている所以

当然ここは、recC(1004) は変異していない箇所なので

Chi と 11 塩基の Chi\* が似ている所以なのではないかと考えられる。

この9アミノ酸に注目して、これらの保存率を数種で比較した。

	RecC identity	特定部位保存度 (recC1004 変異部位)	認識配列
<i>E. coli</i> K-12	100%	9/9 アミノ酸	gctggtgg (nomal chi)
<i>E. coli</i> O-157	100%	9/9 アミノ酸	gctggtgg (nomal chi)
<i>E. coli</i> RecC1004	99.6%	0/9 アミノ酸	gctggtgctcg
<i>Salmonella</i>	88.6%	9/9 アミノ酸	?
<i>H. influenzae</i>	37.1%	3/9 アミノ酸	gNtggtgg, g[c/g]tggagg
<i>Buchnera. sp</i>	28.0%	1/9 アミノ酸	?

しかしながらこの結果が大腸菌とインフルエンザ菌のアミノ酸の違いは $\chi$ 配列の認識配列を甘くしている原因となる部位であるのか、ただの変異の痕跡であるのかを区別できないという理由から、決定的証拠にはなり得ない。

他種に関しては、サルモネラ菌の場合（他の領域も相同性が非常に高いが）この9アミノ酸が完全に保存されていることから、 $\chi$ 配列は大腸菌のものと全く同じであるということが十分考えられる。相同性がそれほど高くない *Buchnera sp.* においても、9アミノ酸の後半部分のうち（インフルエンザ菌ほど多くはないが）10アミノ酸ほどは綺麗に保存されている。この *Buchnera sp.* は他の遺伝子は大腸菌と非常に相同性があるが、recC 遺伝子に関しては相同性が低いという報告がなされている [25]。recA 遺伝子がないため相同組換えの機構がなくなり、相同組換えに必須の RecC に変異がたくさん生じたのではないかとされている。こういった特徴を持っていながらも、この9アミノ酸の後ろの部位だけは多少保存されていることから、recA 遺伝子を失って組換え機構が働かなくなったのが比較的最近であるのかもしれない。

## 5 考察

### ◆ overrepresentation

2コドンを中心に前後にどのような塩基が来る可能性が高いのかを調べたマルコフ解析の結果において、ほとんどの読み枠においても、O/E 値で 1.0 付近を示していることから、 $\chi$ 配列のみに圧力をかけて高出現頻度にするといった進化的な選択圧がかかっていなかったことが分かる。

大腸菌 $\chi$ 配列の読み枠の約7割を占める Leu-Val で解析を行なった図1の結果より、コドン [ctg]-[gtg]の前後に特定の塩基が好まれる傾向がない。 $\chi$ 配列は遺伝子領域に多く存在することから（遺伝子領域はグアニンに富む）前後にグアニンが偶然に位置することにより $\chi$ 配列となったと考えられる。大腸菌の $\chi$ 配列のみが異常な高出現頻度を示しているが、これに関してもこれまで2コドン頻度では説明がつかなかったが、今回の解析で、 $\chi$ 配列の高出現頻度は進化的圧力がかかっていないということが数値的に証明できた。故に大腸菌 RecBCD はもともとコドン頻度の影響で数が多かった g-ctg-gt-g を $\chi$ 配列として利用した可能性が高い。このことは組換え酵素が認識配列を選択している、という仮説を裏付けている。

これまで $\chi$ 配列の過剰出現に関して、1 or 2 コドン頻度などで説明が出来なかった理由として、大腸菌ゲノムでの低出現頻度で知られる "ctag" [3] の影響ではないかと考えられる。Leu-Val の2コドンが連続する際に、Leu をコードするアミノ酸の中で最も頻繁に使用されている [ctg] コドンであるが、このあとにアデニンが来ることは、"ctag" というオリゴヌクレオチドを作り出してしまうため、アデニン塩基が次に来ることを避ける何らかの選択圧が働いているのかもしれない。つまり単純に [ctg] コドンの後ろに来る塩基の確立が 1/4 ではなく、1/3 に近い値のため ([ctg]-A 以外の) [ctg]-[gtg] が多くなってしまったのではないだろうか。つまりこれまで2コドン頻度などで説明ができなかった理由としては "ctag" のオリゴヌクレオチドの並びを避けるため、見かけ上 $\chi$ 配列を増やす選択圧であるかのように思われたと考えられる。



他種ゲノムに関しては、インフルエンザ菌の場合は2コドン頻度で出現頻度の説明がほぼ可能である[17]。前述した“ctag”を避ける傾向は大腸菌のみの特徴であるので、インフルエンザ菌では2コドンで説明が可能なのかもしれない。また枯草菌では、 $\chi$  配列の出現頻度はそれほど高くない[5]。

#### ◆ chi recognition

枯草菌の $\chi$ 配列が5塩基である理由としては、外部からゲノムを取り込む必要があるからではないかと考えられる。外部ゲノムを取り込むという特性のある枯草菌ではある程度自分に似たゲノムで、しかしながら全く似通っていないゲノムはエキソ活性で排除するという機構としては、5塩基程度の認識配列でフィルターをかけるのがちょうど良かったのかもしれない。例えば納豆菌では、同じバチラス族でも外部からゲノムを取り込む機構を持っていないと聞くので、納豆菌に $\chi$ 配列があるとすると、枯草菌の $\chi$ 配列の5塩基よりもっと長い配列である可能性が非常に高いのではないだろうか。

#### ◆ chi sequences at AT-rich species

GC含有量が高い種では $\chi$ 配列がG-richであることは当然の結果だが、AT含有量が高い種では「数が多いことが重要」であるならば、何故 $\chi$ 配列はAT-richではないのだろうか？

当然AT含有量に富む種では多用されるコドンもAT-richであり、コドンの3番目の揺らぎがATに富んでいるという顕著な傾向があるので、多用される配列（例えばコドンなど）を利用して組換え促進のシグナル配列（= $\chi$ 配列）としているのであれば、ATに富む種では、 $\chi$ 配列はAT-richであっても不思議はないはずである。

それでも現在 $\chi$ 配列が実験的に証明されている数種に共通な $\chi$ 配列の特性は「グアニンリッチであること」[17]なので、 $\chi$ 配列がグアニンを多く使わなくてはならない何かしらの理由が存在するはずである。故に、AT含有量が高い種では認識配列を甘くしたり（インフルエンザ菌）、塩基数を減らしてまでも（枯草菌・ラクトコッカス菌） $\chi$ 配列をG-richな配列としておくのではないだろうか。AT含有量の高い種の場合、グアニンが使われているコドンの中で多いものを利用しているというよりもむしろ、 $\chi$ 配列の認識配列を甘くしたり、塩基数を減らすといった手段をとっているようだ。（data not shown）

この理由として考えられるのがRecAとのaffinity[23]などが考えられる。もしかすると結局複製方向に向くためにグアニンを利用している可能性も捨てきれないが、 $\chi$ 配列に方向性が見られないインフルエンザ菌ではこれは説明できないはずである。

つまり、まとめると以下のように考察できる。

- ゲノムのどこの部位でも組換えられるように、数が多いことは重要である。
- $\chi$ 配列はG-richでなければならない理由がある。（← RecA affinity?）
- だからこそ、AT含有量の高い種では認識配列を甘くして（短くして）まで $\chi$ 配列がG (C)-richに成るようにしているのではないか？

## 謝辞

研究方針の相談など頻繁にアドバイスをいただいた中山洋一講師には深く感謝しております。

ネットワークがダウンした際に、親切にも自宅のADSL回線を快く使用させていただいた鷺尾尊規講師に心からお礼を申し上げますと共に、感謝の念でいっぱいです。またネットワーク復旧に尽力を尽くしていただいたITCの山方さんにもこの場を借りてお礼申し上げます。

技術的なサポートしていただいた最上丈仁氏、荒川和晴氏、四津谷 健志氏、久間大輔氏には大変お世話になりました。

最後に金井助教授、内藤講師をはじめとするIABのスタッフの方々、そして研究の場を与えてくださった富田 勝教授に多大なるお礼の言葉と気持ちを伝えさせていただきたいと思います。

## 参考文献

- [1] Kowalczykowski, S. C., Dixon, D. A., Eggleston, A. K., Lauder, S. D., Rehrauer, W. M.; **Biochemistry of homologous recombination in Escherichia coli.**, (1994) *Microbiol.* 58, 401-465.
- [2] Arnold, D. and Kowalczykowski, S. C.; **RecBCD Helicase / Nuclease**, (1999) <http://www.els.net/elsonline/html/A0000586.html>
- [3] Blattner, F. R., Plankett III, G., Bloch, C. A., Perna, N. T., Burland, V., Riley, M., Collado-Vides, J., Glasner, J. D., Rode, C. K., Mayhew, G. F., Gregor, J., Davis, N. W., Kirkpatrick, H. A., Goeden, M. A., Rose, D. J., Mau, B. and Shao, Y.; **The Complete Genome Sequence of Escherichia coli K-12**, (1997) *Science* 277, 1453-1462.
- [4] Sourice, S., Biaudet, V., Karoui, M. E., Ehrlich, D. and Gruss, A.; **Identification of the Chi site of Haemophilus influenzae as several sequences related to the Escherichia coli Chi site**, (1998) *Molecular Microbiology* 27, 1021-1029.
- [5] Chedin, F., Noirot, P., Biaudet, V. and Ehrlich, D.; **A five-nucleotide sequence protects DNA from exonucleolytic degradation by AddAB, the RecBCD analogue of Bacillus subtilis.**, (1998) *Molecular Microbiology* 29, 1369-1377.
- [6] Chedin, F., Ehrlich, S. D., Kowalczykowski, S. C.; **The Bacillus subtilis AddAB helicase/nuclease is regulated by its cognate Chi sequence in vitro.**, (2000) *J. Mol. Biol.* 298, 7-20.
- [7] Biswas, I., Maguin, E., Ehrlich, S. D. and Gruss, A.; **A 7-base pair sequence protects DNA from exonucleolytic degradation in Lactococcus lactis.**, (1995) *Proc Natl Acad Sci USA*, 92, 2244-2248.
- [8] Karoui, M., Ehrlich, D. and Gruss, A.; **Identification of the lactococcal exonuclease / recombinase and its modulation putative Chi sequence.**, (1998) *Proc Natl Acad Sci USA*, 95, 626-631.
- [9] Biaudet, V., El Karoui, M., Gruss, A.; **Codon usage can explain GT-rich islands surrounding Chi sites on the Escherichia coli genome.**, (1998) *Mol. Microbiol.* 29, 666-669.
- [10] Colbert, T., Taylor, A. F., Smith, G. R.; **Genomics, Chi sites and codons: 'islands of preferred DNA pairing' are oceans of ORFs.**, (1998) *Trends Genet.* 14, 485-488.
- [11] Kuzminov, A.; **Collapse and repair of replication fork in Escherichia coli**, (1995) *Mol Microbiol.* 16, 373-84.
- [12] Cox, M. M.; **A broadening view of recombinational DNA repair in bacteria.**, (1998) *Genes Cells* 3, 65-78.

- [13] Reina, U., Nakayama, Y., Arakawa, K., Tomita, M.; **Orientation of Chi sequence is a general tendency of G-rich oligomers**, (2000) *Gene*, 259, 207-215.
- [14] Lobry, J. R.; **Asymmetric Substitution Patterns in the Two DNA Strands of Bacteria**, (1998) *Molecular Biology and Evolution* 13, 660-665.
- [15] Bell, S.J., Chow, Y. C., Ho, J. Y. K. and Forsdyke, D. R.; **Correlation of Chi orientation with transcription indicates a fundamental relationship between recombination and transcription.**, (1998) *Gene* 216, 285-292.
- [16] Lao, P. J., Forsdyke, D. R.; **Crossover hot-spot instigator (Chi) sequences in Escherichia coli occupy distinct recombination/transcription islands.**, (2000) *Gene* 243, 47-57.
- [17] El Karoui, M., Biaudet, V., Schbath, S., Gruss, A.; **Characteristics of Chi distribution on different bacterial genomes.**, (1999) *Res. Microbiol.* 150, 579-587.
- [18] Handa, N., Ohashi, S., Kusano, K., Kobayashi, I.; **chi<sup>\*</sup>, a chi-relatex 11-mer sequence partially active in an E. coli recC<sup>\*</sup> strain**, (1997) *Genes to Cells* 2, 525-536.
- [19] Arnold, D.A., Handa, N., Kobayashi, I., Kowalczykowski S.C.; **A novel, 11 nucleotide variant of chi, chi<sup>\*</sup>: one of a class of sequences defining the Escherichia coli recombination hotspot chi.**,(2000) *J Mol Biol.*, 300(3):469-79.
- [20] Smith, G. R., Amundsen, S. K., Dabert, P. and Taylor A. F.; **The initiation and control of homologous recombination in Escherichia coli.**, (1995) *Philos Trans R Soc Lond B Biol Sci.*, 347, 13-20.
- [21] Dixon, D. A. and Kowalczykowski, S. C.; **Homologous Pairing In Vivo Stimulated by the Recombination Hotspot, Chi.**, (1991) *Cell*, 66, 361-371.
- [22] Eggleston, A. K., West, S. C.; **Recombination initiation: easy as A, B, C, D... chi?** , (1997) *Curr. Biol.* 7, R745-749.
- [23] Tracy, R.B. and Kowalczykowski, S.C.; **In vitro selection of preferred DNA pairing sequences by the Escherichia coli RecA protein.** (1996) *Genes dev.* 10, 1890-1903.
- [24] Tracy, R. B., Chedin, F. and Kowalczykowski, S. C.; **The Recombination Hot Spot Chi Is Embedded within Islands of Preferred DNA Pairing Sequences in the E.coli Genome**, (1997) *Cell* 90, 205-206.
- [25] Shigenobu, S., Watanabe, H., Sakaki, Y., Ishikawa, H.; **Accumulation of Species-Specific Amino Acid Replacements That Cause Loss of Particular Protein Functions in Buchnera, an Endocellular Bacterial Symbiont.**, (2001)*J Mol Evol* 53:377-386.
- [26] Churchill, J. and Kowalczykowski, S., C.; **Identification of the RecA Protein-loading Domain of RecBCD Enzyme**, (2000) *J. Mol. Biol.* 297, 537-542.
- [27] Kunst F. et al.; **The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.**, (1997) *Nature*, 390, 249-256.

- [28] Fleischmann, R.D., Adams, M.D., White, O., Clayton, R.A., Kirkness, E.F., Kerlavage, A.R., Bult, C.J., Tomb, J.F., Dougherty, B.A., Merrick, J.M. and et al.; **Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd.**, (1995) *Science* 269, 496-512.