

東京大学大学院新領域創成科学研究科  
情報生命科学専攻

平成 17 年度

修士論文

キメラ化 Sip4 転写因子による下流標的遺伝子の網羅的同定

2006 年 3 月提出

指導教員 伊藤 隆司 教授

46909 谷 祐治

## 目次

### 1. 序論

### 2. 材料と方法

#### 2-1 使用菌株

#### 2-2 使用プライマー

#### 2-3 使用培地

#### 2-4 菌株の維持・取り扱い

#### 2-5 pRS415chimeraZEO-Sip の作成

#### 2-6 RT-PCR

#### 2-7 マイクロアレイによる発現解析

#### 2-8 発現プロファイルを使ったコンセンサス配列抽出プログラム

### 3. 結果

#### 3.1 キメラ化 Sip4 のデザイン

#### 3.2 キメラ化 Sip4 発現株の作成

#### 3.3 キメラ化 Sip4 発現株のマイクロアレイによる発現遺伝子解析

#### 3.4 標的遺伝子濃縮の試み

#### 3.5 結合配列解析の試み

### 4. 考察

5. 参考文献

6. 謝辞

7. 発表状況

8. 表と図

# 1. 序 論

近年の急速なゲノム解析の発展により、多くの生物において全ゲノム塩基配列が決定され、多数の新規遺伝子が発見された。しかしながら、新規遺伝子の相当部分は構造だけからでは機能の推測すら付かないもので占められており、それらの中には種を超えて広く保存されている物も少なくない。こうした遺伝子は細胞機能の基本的な部分に関与していると推測されるが、既存の方法論では捉えきれなかったか、あるいは従来までの研究の視野に入っていなかった未知の生物機能に関与する物である可能性も高い。したがって、それらの機能を明らかにすることは今後の重要な課題である。

新規遺伝子の機能解析のひとつの鍵は、遺伝子制御ネットワークを解明し、そこにその遺伝子を位置づけることであると考えられる。例えば、ある新規遺伝子が機能既知の転写因子の制御を受けていることがわかれば、その制御関係情報は機能未知遺伝子の機能推定の重要な手がかりになる。逆に機能未知の転写因子も数多く存在するが、その標的遺伝子群を同定することが出来て、それらの中に機能がよくわかった遺伝子があれば、そこから逆に転写因子の機能を推測することができる(Fig.1)。

遺伝子の標的遺伝子群の検索のアプローチとして、転写因子変異株のトランスクリプトーム解析やゲノムワイドのクロマチン免疫沈降(ChIP)が行われている。しかし、多くの転写因子は、その活性化に様々な生育条件等を反映した上流からのシグナル伝達を必要とする。したがって、活性化刺激を先験的に知ることが出来ない新規転写因子では、その破壊株や単純な過剰発現株を用いた標的遺伝子探索には限界がある。

この問題を克服するために、当研究室で転写因子のキメラ化による恒常活性化戦略(Chimerization-mediated Activation: CMA)が、出芽酵母で最大の転写因子ファミリーである Zn<sub>2</sub>Cys<sub>6</sub> 型転写因子をモデルに開発された。この方法は、Zn<sub>2</sub>Cys<sub>6</sub> 型転写因子の DNA 結合ドメインよりも C 末端側の部分を恒常的に転写を活性化する VP16 で置換することでキメラ化し、その標的遺伝子の発現を恒常的に誘導しようとする方法である(Fig.2)。この方法により、多剤耐性を制御する転写因子 Pdr1 など、それまで単純な過剰発現や遺伝子破壊では下流標的遺伝子を同定出来なかった転写因子について、既知の標的遺伝子の発現を誘導することが可能であることが示されている。しかし、この方法では直接の標的遺伝子とともに多くの 2 次的な標的遺伝子の発現が誘導され、候補遺伝子群の中から直接の標的遺伝子を抽出する手法が必要となっている。

真の標的遺伝子を抽出する方法としては、現在、標的遺伝子の上流配列に存在する転写因子結合配列や共通配列を抽出するアプローチが広く行われている。**Zn<sub>2</sub>Cys<sub>6</sub>**型転写因子の結合配列は、保存された2つの領域とその間をつなぐ保存性の低いスペーサー領域から構成されていることが知られている。こうした形状の結合配列を抽出するための優れたプログラムとしては、“**Dyad analysis**”が知られている。この方法は、生物学的知見を考慮せずに、比較的単純な統計学的手法のみを用いて結合配列を抽出する方法である。

本研究では、**Sip4** 転写因子を対象に取り上げキメラ転写因子作成とマイクロアレイによる発現解析を行うことにより下流標的遺伝子候補の同定を行い、更に“**Dyad analysis**”に近年急激に蓄積されてきた種々の生物学的知見を組み入れることによって開発したより精度の高い結合配列抽出方法を適用することを試みた。

**Sip4** の **coiled-coil** ドメインを **COIL** プログラムで予測し、それより下流領域を **λcI** リプレッサーと **VP16** で置換したキメラ転写因子を作成した。これを発現させた細胞を用いてマイクロアレイ実験を行ない、対照細胞に比較して発現が4倍以上上昇した遺伝子を **42** 個見出した。これらの遺伝子群を **GO term finder** で解析すると、糖代謝関連の遺伝子が有意に濃縮されていることが分った。また、それらの上流配列を解析したところ **20** 遺伝子が **CSRE** 様配列を持っていることも分った。残りの **22** 遺伝子は **CSRE** を持たず、非特異的、もしくは二次的な標的遺伝子であると考えることが出来る。そのため、これらの **22** 遺伝子を排除し、標的遺伝子のみを抽出する手法の開発を行った。**Sip4** をベンチマークとし、その他のキメラ化転写因子に関しても適用し、評価を行った。

## 2. 材料と方法

### 2-1 使用菌株

出芽酵母は FY1679-28C (*MATa*, *ura3-52*, *leu2-Δ1*, *trp1-Δ63*, *his3-Δ200*, *GAL2+*)を、大腸菌は DH5α を使用した。

### 2-2 使用培地

本研究において使用した培地(液体)の組成を以下に示す。固体培地として使用する場合には 2 % (w/v)の Bacto Agar (BD)を加えた。すべて調整後、加圧滅菌 (121°C, 20 分)してから使用した。液体培地は原則的に 500 ml ずつ作成し、室温保存した。固体培地は使用する量に合わせて作成し、火炎滅菌下で注ぎ分け封をして室温暗所にて保存した。

出芽酵母(*S. cerevisiae*)用培地

#### 1) 完全培地(YPAD)

Yeast Extract	10.0 g
BactoPeptone	20.0 g
Adenine	0.1 g
Dextrose	20.0 g
DW(Distilled water)	1000 ml

必要に応じて Zeocin (最終濃度 75 μg/ml) を加えた YPAD-Zeo 寒天培地を作成した。

#### 2) ラフィノース培地(YPR)

Yeast Extract	10.0 g
BactoPeptone	20.0 g
Raffinose	20.0 g
DW	1000 ml

大腸菌(*E. coli*)用培地

1) LB amp 選択培地(LBA)

Bactopeptone	10 g
Yeast Extract	5 g
NaCl	5 g
Ampicilin	50 mg
DW	1000 ml

2) 2 x YT 培地

Bactopeptone	16 g
Yeast Extract	10 g
NaCl	5 g
Ampicilin	50 mg
DW	1000 ml

## 2-3 使用プライマー

本研究で使用した合成オリゴヌクレオチドプライマーはまとめて表.2 に示す。

## 2-4 菌株の維持・取り扱い

### 2-4-1 出芽酵母(*S. cerevisiae*)

原則として、長期間の保存が必要な場合は 30%グリセロールに等量の YPAD によって培養した培地を等量加え、 $-80^{\circ}\text{C}$ のディープフリーザーの中でストックとして維持した。短期間の保存の場合は、 $4^{\circ}\text{C}$ においてプレートをテープで封をして保存した。また、植菌などの操作は全て火炎滅菌下で行った。

### 2-4-2 大腸菌(*E. coli*)

出芽酵母と同様に取扱いは全て火炎滅菌下で行った。

## 2-5 pRS415chimeraZEO-Sip の作成

出芽酵母 *GAL1* プロモーター、 $\lambda$  ファージ *cI* リプレッサー、単純ヘルペスウ

イルス VP16、Strep-His-Flag タンデムタグ、出芽酵母 *ADH1* ターミネーター及び Zeocin 耐性遺伝子を pRS415 (Stratagene) 上に組み込んだプラスミド pRS415chimeraZEO (Fig. 1?) に、以下に述べるギャップ修復法を用いて Sip4 の DNA 結合領域をコードする DNA 断片をクローニングした。

#### 2-5-1 Sip4 の DNA 結合領域の増幅

Sip4 の翻訳開始点から coiled-coil 領域をコードする配列を含む領域 (822 bp) のクローニングを行った。coiled-coil 領域の予測にはインターネット上のサイトプログラム COILS-Prediction of Coiled Coil Regions in Proteins を用いた ([http://www.ch.embnet.org/software/COILS\\_form.html](http://www.ch.embnet.org/software/COILS_form.html))。

##### 2-5-1-1 翻訳開始点から coil-coil 領域を含む領域の PCR による増幅 (FIG. 3 ①)

GeneAmp PCR System9700 (Applied Biosystems) を用い、以下の反応組成液、PCR プログラムを用いて PCR 反応を行った。

##### <反応液の組成>

10x PCR Buffer	10.0 $\mu$ l
dNTP mix	10.0 $\mu$ l
Template DNA*1	2.0 $\mu$ l
Ex-Taq DNA polymerase	0.5 $\mu$ l
Sip4-CF1 primer(10pmol)	3.0 $\mu$ l
Sip4-CF2 primer(10pmol)	3.0 $\mu$ l
DW	71.5 $\mu$ l

##### \*1 Template DNA の調製

YPAD プレート上で培養したコロニー (FY1679-28C) を爪楊枝で少量掻き取り、2 x Zymolase 溶液 40  $\mu$ l に懸濁する

↓

37°C 30分

↓

室温 15000 rpm で1分遠心する

↓

沈殿物に DW 40  $\mu$ l を加え、ボルテックスを用いて懸濁する

↓

室温 15000 rpm で1分遠心する

↓



沈殿物に以下の組成(\*2)の溶液を加え、タッピングにより懸濁する

↓

55°C 1時間

↓

TE 40  $\mu$ l、フェノール/クロロホルム 50  $\mu$ lを加え、ボルテックスにより混和する

↓

室温 15000 rpm で3分遠心する

↓

上清を新しいチューブに回収し、フェノール/クロロホルム 50  $\mu$ lを加え、ボルテックスにより混和する

↓

室温 15000 rpm で3分遠心する

↓

上清を新しいチューブに回収し、クロロホルム 50  $\mu$ lを加え、ボルテックスにより混和する

↓

室温 15000 rpm で3分遠心する

↓

上清を鋳型として使用する

(\*2)

TE	9.0 $\mu$ l
10% SDS	1.0 $\mu$ l
5M NaCl	1.0 $\mu$ l
Protease K 溶液	1.0 $\mu$ l

#### PCR プログラム

95°C	3:00
95°C	0:15
55°C	0:15
72°C	1:00 25 サイクル
72°C	5:00

PCR 産物は1% アガロース TAE ゲルを用いた電気泳動(100 V)により目的の長さ (822 bp) の断片が生成されているかを確認した。

### 2-5-1-2 PCR 産物のエタノール沈殿

上記の反応で得られた目的の長さの PCR 産物を以下のプロトコールでエタノール沈殿を行った

PCR 産物にエタノール 250  $\mu$ l、酢酸ナトリウム(pH 5.2)10  $\mu$ l を加える

↓

4°C 15000 rpm 9 分遠心する

↓

上清を捨て、沈殿物に 75%エタノールを 250  $\mu$ l 加えてリンス  
エタノールを取り除き沈殿物をおる程度乾燥させる

↓

TE 10  $\mu$ l に溶解させる

### 2-5-2 *GAL1* プロモーター/結合ドメイン・リンカー断片の作成 (Fig.3 ②)

制限酵素処理した pRS415chimeraZEO (2-5-4 参照) を鋳型にして、下記の反応組成液を用いて 2-5-1-1 と同様に PCR 反応を行い、目的の長さの断片(645 bp)が増幅されていることを確認した後 2-5-1-2 と同様にエタノール沈殿を行った。

10x PCR Buffer	5.0 $\mu$ l
dNTP mix	5.0 $\mu$ l
Templata DNA*3	2.0 $\mu$ l
Ex-Taq DNA polymerase	0.5 $\mu$ l
pCUF primer(10 $\mu$ mol)	1.5 $\mu$ l
Sip4-PR2(10 $\mu$ mol)	1.5 $\mu$ l
DW	35.5 $\mu$ l

### 2-5-3 結合ドメイン/活性化ドメイン・リンカー断片の作成 (Fig.3 ③)

制限酵素処理した pRS415chimeraZEO (2-5-4 参照) を鋳型にして、下記の反応組成液を用いて 2-5-1-1 と同様に PCR 反応を行い、ゲル電気泳動により目的のサイズの断片 (540 bp) を確認した後 2-5-1-2 と同様にエタノール沈殿を行った。

10x PCR Buffer	5.0 $\mu$ l
dNTP mix	5.0 $\mu$ l

Templata DNA*4	2.0 $\mu$ l
Ex-Taq DNA polymerase	0.5 $\mu$ l
Sip4-LambdaF primer(10pmol)	1.5 $\mu$ l
pVP16 primer(10pmol)	1.5 $\mu$ l
DW	35.5 $\mu$ l

#### 2-5-4 pRS415chimeraZEO の制限酵素処理

(Fig.3 ④)

pRS415chimeraZEO を以下の反応溶液で制限酵素処理を行った。

10×H buffer	2.0 $\mu$ l
pRS415chimeraZEO	2.0 $\mu$ l
DW	15.0 $\mu$ l
BglII	0.5 $\mu$ l
MluI	0.5 $\mu$ l

37°C 1時間

↓

プラスミドが切断されているか確認するためにゲル電気泳動を行う

↓

フェノール-クロロホルム 20  $\mu$ l を加え、ボルテックスにより混和する

↓

室温 15000 rpm で3分遠心する

↓

上清に上清の2.5倍量のエタノール、10分の1量の酢酸ナトリウム (pH 5.2) を加える

↓

4°C 15000 rpm 9分 遠心

↓

100  $\mu$ l TE に溶解する

#### 2-5-5 ギャップ修復クローニング法

以下のプロトコールのギャップ修復クローニング法によりコンストラクトを作成した。

##### 2-5-5-1 トランスフォーメーション

YPAD プレートからループで Y1679-28C のコロニーを掻き取り、10 ml YPAD  
で 30°C 一晩振とう培養

↓

YPAD で 100 倍希釈

↓

2 時間振とう培養

↓

2 ml をチューブにとり室温 15000 rpm で 10 秒遠心し集菌

↓

滅菌水で菌体を洗浄

↓

溶液 A(下記)及び 2-5-1, 2-5-2, 2-5-3, 2-5-4 で得られた DNA 断片を加え、  
ボルテックスにより菌体を懸濁する

↓

30°C 30 分 培養器で振とうする

↓

DMSO 36  $\mu$ l を加えボルテックスにより混和し、42°C で 20 分静置

↓

15000 rpm で 10 秒遠心し、集菌し、YPAD 750  $\mu$ l を加え 30°C で 2 時間培養を  
行う(回復培養)

↓

15000 rpm で 10 秒遠心し、集菌する

↓

滅菌水 100  $\mu$ l を加え、ボルテックスにより懸濁する

↓

懸濁液を YPAD-Zeo plate に塗布し、30°C で二日間培養する

#### 溶液 A

50% PEG	240 $\mu$ l
1M 酢酸リチウム	36 $\mu$ l
DW	60 $\mu$ l
キャリアー DNA*	5 $\mu$ l

\*10 mg/ml を使用直前に熱処理、急冷し使用した。

#### 2-5-5-2 酵母からのプラスミド回収

SV Miniprep (Promega)を用いて酵母からプラスミドを回収した。

2-5-5-1により得られたプレート上のコロニーを全て掻き取り、下記の溶液 A を加え、ボルテックスにより懸濁する

↓

37°C 1時間静置する

↓

Solution 1 を 150  $\mu$ l 加える

↓

Solution 2 を 250  $\mu$ l 加え、転倒混和する

↓

Alkaline protease を 10  $\mu$ l 加える

↓

37°C 30分 静置する

↓

Solution 3 を 350  $\mu$ l 加え転倒混和する

↓

クロロホルム 100  $\mu$ l を加え、ボルテックスにより混和する

↓

室温 15000 rpm 5分 遠心する

↓

上清をキット付属のカラムに乗せ、室温 15000 rpm 1分遠心する

↓

カラムに 750  $\mu$ l の wash バッファーに乗せ、室温 15000 rpm 1分遠心する

↓

カラムに 250  $\mu$ l の wash バッファーに乗せ、室温 15000 rpm 2分遠心する

↓

100  $\mu$ l の TE を加え、室温 15000 rpm 1分遠心

#### 溶液 A

10 x Zymolase	10.0 $\mu$ l
グリセロール	12.5 $\mu$ l
B-メルカプトエタノール	0.5 $\mu$ l

#### 2-5-5-3 大腸菌のトランスフォーメーション

2-5-5-2 で得られたプラスミド溶液は非常に濃度が低く、かつ様々なコンスト

ラクトが混じっている事が考えられるため、大腸菌に対するトランスフォーメーションを行った。

大腸菌 DH5 $\alpha$  コンピテントセル 50  $\mu$ l を室温で融解させる

↓

2-5-5-2 で得られた溶液 5  $\mu$ l を加え、氷上で 20 分静置する

↓

42 $^{\circ}$ Cの恒温槽に 30 秒 静置

↓

LBA プレートに塗布し、37 $^{\circ}$ Cで二日間培養する

2-5-5-4 大腸菌からのプラスミド回収

2-5-5-3 により得られた大腸菌を爪楊枝で少量掻き取り、2 x YT 2 ml に植菌する。

↓

37 $^{\circ}$ C 6時間 振とう培養

↓

室温 15000 rpm 3分 遠心し、集菌する

↓

solution1 を 100 $\mu$ l 加え、ボルテックスにより懸濁する

↓

Solution2 を 200 $\mu$ l 加え、転倒混和する

↓

Solution3 を 150 $\mu$ l 加え、転倒混和する

↓

フェノール/クロロホルム 150 $\mu$ l を加え、ボルテックスにより混和する

↓

室温 15000rpm 3分遠心し上清を回収する

↓

エタノール 400  $\mu$ l, sodium acetate (pH 5.2)

40 $\mu$ l 加え、ボルテックスにより混和する

↓

4 $^{\circ}$ C 15000rpm 9分 遠心する

↓

上清を捨て、75%エタノール 500  $\mu$ l を加え転倒混和する

↓  
ペレットに TE 100  $\mu$ l , 1mg/l RNase 1  $\mu$ l を加え、ボルテックスにより溶解させる  
↓  
37°C 30分 静置  
↓  
20% PEG /2.5M NaCl 100  $\mu$ l を加える  
↓  
4°C 15000 rpm 9分 遠心  
↓  
ペレットに 75% エタノール 300 $\mu$ l を加えボルテックスにより混和する  
↓  
4°C 15000 rpm 9分 遠心  
↓  
TE 50  $\mu$ l に溶解する

### Solution1

50 mM グルコース  
25 mM Tris-HCl(pH 8.0)  
10 mM EDTA

### Solution2

0.2 N 水酸化ナトリウム  
1% SDS

### Solution3

300mM 酢酸カリウム  
115mM 酢酸

### フェノール/クロロホルム

中性フェノール 25 ml  
クロロホルム 25 ml

TE を重層し、pH.8.0 を確認する

2-5-5-5 制限酵素によるインサートの確認

得られたプラスミドを制限酵素 **BamH I** ,**Sph I** で消化した。

10×H buffer	2.0 µl		
大腸菌により得られたプラスミド	2.0 µl		
DW	15.0 µl		
BamH I	0.5 µl	37°C	1時間 静置する
Sph I	0.5 µl	↓	
			1% アガロース TAE ゲル

を用いた電気泳動(100 V)を行い、目的の長さのバンド(1090 bp, 679 bp)を確認する。

#### 2-5-6 酵母へのトランスフォーメーション

制限酵素処理により、目的の長さのバンドが見られたプラスミドを 2-5-5-1 と同様のプロトコールにより酵母にトランスフォームした。

#### 2-5-7 キメラ化 Sip4 の発現確認(western blotting 法)

以下のプロトコールに従ってキメラ化 Sip4 の発現を確認した。

コロニーをループで掻き取り、YPR で振とう培養(オーバーナイト)

↓

全量 50 ml になるように 100 倍希釈

↓

バッフル付き三角フラスコで 30°C 4 時間 振とう培養

↓

20% ガラクトース 5 ml(最終濃度 2 %) を添加

↓

2 時間 振とう培養

↓

培養液 1 ml をエッペンに移す

↓

ベータアルカリを加え(下記)、ボルテックスにより混和する

↓

氷上 5 分

↓

30% TCA 130 µl を加える

↓

氷上 5 分



↓  
4°C 15000 rpm 1分 遠心する  
↓  
上清を捨て、4°Cアセトン 500  $\mu$ l を加える  
↓  
ソニケーションにより沈殿を破碎する  
↓  
4°C 15000 rpm 10秒 遠心する  
↓  
完全にアセトンを取り除き乾燥させる  
↓  
サンプルバッファー（下記） 100  $\mu$ l を加え、沈殿を懸濁する  
↓  
95°C 5分  
↓  
室温 15000 rpm 3分遠心を行う  
↓  
10%SDS-ポリアクリルアミドゲル(組成 下記)に 20  $\mu$ l をロードし、電圧：300 V、電流：30 mA、泳動時間：1時間の条件で泳動を行った  
↓  
泳動後、セミドライブロッティング装置とブロッティングバッファー（下記）を用いニトロセルロース膜(MILIPORE)へとブロッティングを行った(電圧：20 V、電流：300 mA、一時間)  
↓  
メンブレンを DW で2~3回洗う  
↓  
Ponseau-S を適当量加える  
↓  
DW で2~3回洗う  
↓  
ブロッキング溶液 30ml を加え、30分 振とうする  
ブロッキング溶液 10ml に抗FLAG抗体 1  $\mu$ l を加え、メンブレンにかける  
↓  
室温 30分 静置  
↓  
TPBS 250 ml を用いて3回メンブレンを洗う

↓  
二次抗体として TPBS 2 ml に二次抗体(抗マウス Ig 抗体) 1 μl を加えメンブレンにかける

↓  
(TPBS で洗う+TPBS 溶液中で振とう 15 分) を数回繰り返す

↓  
ECL ウェスタンブロットティング検出システム (Amersham) でバンドを発光、FUJI Film Las-3000 にて検出

↓  
バンドの確認(37.076kDa)

<u>ベータアルカリ</u>	
メルカプトエタノール	11.8 ml
10N NaOH	30.0 ml
DW	108.2 ml

<u>ランニングゲル</u>	
0.75 M Tris-HCl(PH.8.8)	5.0 ml
30% AA	6.7 ml
10% SDS	0.4 ml
DW	2.9 ml

+	
TEMED	10 μl
10% APS	75 μl

<u>スタッキングゲル</u>	
0.25M Tris-HCl(PH.6.8)	3.00 ml
30% AA	0.72 ml
10% SDS	0.12 ml
DW	2.30 ml

+	
TEMED	6μl
10% APS	60μl

<u>TPBS</u>	
NaCl	8.0 g
Na <sub>2</sub> HPO <sub>4</sub> · 12H <sub>2</sub> O	2.9 g
KCl	0.2 g

KH <sub>2</sub> PO <sub>4</sub>	0.2 g
DW	fill up to 1000ml

### ブロッキング溶液

TPBS	40ml
スキムミルク	0.8g

### 2X サンプルバッファー

0.5M Tris-HCl(pH.6.8)	2.0 ml
10% SDS	4.0 ml
β-メルカプトエタノール	1.2 ml
グリセロール	2.0 ml
DW	0.8 ml
BPB	少量

### 10x 泳動バッファー

Tris	30.3 g
グリシン	143.1 g
SDS	10.0 g
DW	1000 ml

### 10x ブロッキングバッファー

Tris	30.20 g
グリシン	144.14 g
SDS	1.00 g
DW	1000 ml

## **2-6 RT-PCR**

### **2-6-1 RNA 抽出**

以下のプロトコールを用いて RNA 抽出を行った。

RNA 抽出を行う菌体は 2-5-7 と同様に培養し、菌体を 50 ml 全て回収した。  
回収した酵母に DW1000 μl 加え、ボルテックスにより懸濁する

↓

室温 3000 rpm 5 分 遠心し、上清を取り除く

↓

Lysis buffer 600  $\mu$ l (10 mM Tris-HCl (pH7.5), 10 mM EDTA, 0.5%SDS)を加え、ボルテックスにより細胞を懸濁する

↓

水飽和フェノールを加え、65°C 1時間 ボルテックスにより混和を行う

↓

2層に分かれるまで氷上で冷却する

↓

4°C 13000 rpm 3分 遠心

↓

上清に水飽和フェノール 600  $\mu$ l を加えボルテックスにより混和する

↓

4°C 13000 rpm 3分 遠心

↓

上清に、クロロフォルム 600  $\mu$ l を加えボルテックスにより混和する

↓

4°C 13000 rpm 3分 遠心

↓

上清を新しいチューブに移し、3M 酢酸ナトリウム 50  $\mu$ l 及びイソプロパノール 600  $\mu$ l を加えボルテックスにより混和する

↓

4°C 13000 rpm 3分遠心

↓

上清を捨て、75% エタノール 600  $\mu$ l を加えボルテックスにより混和する

↓

4°C 13000 rpm 1分 遠心

↓

上清を捨てる

↓

DEPC 水 80  $\mu$ l に溶解する

↓

10×DNase Buffer 10  $\mu$ l

BSA 10  $\mu$ l

DNaseI 1  $\mu$ l

vanadyl ribonucleoside complex 1~2  $\mu$ l を加える

↓

37°C 30分 静置

↓  
TRIzol 750  $\mu$ l, クロロホルム 200  $\mu$ l を加えボルテックスにより混和する  
↓  
4°C 13000 rpm 3分 遠心を行う  
↓  
上清を新しいチューブに移す  
↓  
等量のクロロホルムを加え、ボルテックスにより混和する  
↓  
4°C 13000 rpm 3分 遠心  
↓  
上清を新しいチューブに移す  
↓  
等量のイソプロパノールを加え、ボルテックスにより混和する  
↓  
4°C 13000 rpm 3分 遠心  
↓  
75% エタノール 600  $\mu$ l を加え、ボルテックスにより混和する  
↓  
4°C 13000 rpm 1分 遠心  
↓  
DEPC 処理水に溶かす  
↓  
分光高度計により RNA 量を測定し、1  $\mu$ g/ $\mu$ l に合わせる

#### Lysis buffer

10mM Tris-HCl (pH7.5)

10mM EDTA

0.5% SDS

#### 2-6-2 cDNA の調製

RNA 2.5  $\mu$ g

Oligo(dT) 1.0  $\mu$ g

DW fill upto 11 $\mu$ l

↓  
70°C 10分 静置する

↓

以下の反応溶液を加える

DEPC 水	2.0 $\mu$ l
5 $\times$ first strand buffer	4.0 $\mu$ l
0.1M DTT	2.0 $\mu$ l
10 $\times$ dNTP	1.0 $\mu$ l
Super script	1.0 $\mu$ l

↓

42°C 50 分

↓

70°C 15 分

↓

TE 180  $\mu$ l を加える

### 2-6-3 RT-PCR

以下の反応組成液、及び反応プログラムを用いて実施した。

PCR 産物を 1% アガロース TAE ゲルを用いた電気泳動(100V)を用いて確認を行った。

<反応液の組成>

dNTP mix	1.0 $\mu$ l
10x PCR Buffer	1.0 $\mu$ l
cDNA	2.0 $\mu$ l
primer mix(10pml)*	1.0 $\mu$ l
H <sub>2</sub> O	5.0 $\mu$ l

\*primer mix は ORF 全長を増幅する “Gene Pairs”(Research Genetics 社)を用いた。

### PCR プログラム

95°C	3:00
95°C	0:15
55°C	0:15
72°C	1:00 25~30 サイクル*
72°C	5:00

\*必要に応じてサイクル数を調整する

## 2-7 マイクロアレイによる発現解析

以下のプロトコールにより解析を進めた。

RNAの調製は2-6-1のプロトコールと同様に行い、最終濃度を5 $\mu$ g/ $\mu$ lにする。

total RNA	50 $\mu$ g
0.5mg/ml oligo(dT)	1 $\mu$ l

上記を混合し、全量を11 $\mu$ lとする

↓

72 $^{\circ}$ C 5分 静置

↓

42 $^{\circ}$ C 5分 静置

↓

以下の反応溶液を加える

5 $\times$ Buffer	4.0 $\mu$ l
0.1 M DTT	2.0 $\mu$ l
10X dNTP(aa)	2.0 $\mu$ l
RNase inhibitor	0.1~0.05 $\mu$ l

↓

42 $^{\circ}$ C 45~60分 静置

↓

Power script を1  $\mu$ l加える

↓

42 $^{\circ}$ C 60分 静置

↓

100 mM EDTA 25  $\mu$ l

1M NaOH 10  $\mu$ l を加える

↓

60 $^{\circ}$ C 60min

↓

1M Tris-HCl (pH7.5)を25  $\mu$ l加える

↓

Microcon30 (MILLIPORE) カラムに乗せる

↓

室温 14000 rpm 5分 遠心する

↓

100  $\mu$ l DW を カラムに乗せ 室温 14000 rpm 5分 遠心する  
(この作業を4回繰り返す)

↓

30  $\mu$ l DW を加え、カラムを新しいチューブに逆さにセットする

↓

室温 5000 rpm 1分 遠心する

↓

0.1 M  $\text{NaHCO}_3$  (pH 9.0) 15  $\mu$ l に懸濁した Cy3 もしくは Cy5 (Amersham)を加え遮光し、室温 で1時間静置する

↓

4M Hydroxylamine HCl を 15  $\mu$ l 加え遮光して室温 15分

↓

PCR purification kit (QIAGEN)で精製を行う

↓

酵母マイクロアレイ (DNA Chip Research Inst.) にハイブリダイズする

↓

65°C 10~20時間 静置

↓

2 x SSC, 0.1% SDS 中でカバーガラスをはずす

↓

2 x SSC, 0.1% SDS でスライドガラスを洗う

↓

1 x SSC でスライドガラスを洗う

↓

0.2 x SSC でスライドガラスを洗う

↓

室温 800 rpm 2分 遠心を行う

↓

Gen Pix 4000A scanner(Axon)にてシグナルをよみとり  
Gene Pix 3.0 software (Axon)により解析を行った。

#### dNTP(aa)

5 mM dATP, dGTP, dCTP

2 mM dTTP

3mM aadUTP (sigma A0140)



### 10 x SSC

NaCl	350.4 g
C <sub>6</sub> H <sub>5</sub> ONa <sub>3</sub>	155.0 g
DW	1000 ml
	HCl で pH を 7.0 に合わせる

## 2-8 発現プロファイルを使った遺伝子濃縮及びコンセンサス配列抽出プログラム

### 出プログラム

開発環境として CPU : AMD Athlon64 3000+(1.8GHz), Memory : 1G OS: WindowsXP , java 開発環境である eclipse に Perl EPIC を導入し、Perl によるプログラム開発を行った。実行環境については上記と同一のパソコンを用い、ActivePerl 上で実行を行った。

実験により得られた遺伝子群から真の標的の濃縮、及び結合配列の抽出を試み、その性能の検証を行った。具体的には以下二つの開発を行った。

- ① 実験により得られた標的候補を様々な条件下の発現パターンに基づいてクラスタリングする事により標的遺伝子を濃縮するプログラム。
- ② 生物学的な知見を統計学的な上流配列に付加する事により、結合配列抽出の精度向上を目指すプログラム。

## 3. 結果

### 3.1 キメラ化 Sip4 のデザイン

Sip4 が属する  $Zn_2Cys_6$  型 zinc フィンガー転写因子ファミリーを代表するのがガラクトース代謝を制御する Gal4 である。Gal4 は真核生物の転写因子の代表として古くから、構造と機能の関連がよく研究されている。その結果、Fig.4 に示すようにこのファミリーのメンバーは、N 末端から順に DNA 結合ドメイン、二量体化促進ドメイン、転写活性化ドメインで構成されることが分かっている。

DNA 結合ドメインは二量体化して標的配列に結合する。このドメインは、DNA を直接認識する zinc フィンガーと、二量体化を司る coiled-coil、および両者をつなぐリンカー領域から成っていて、構造的な保存性が高い。このドメインの C 末端側の境界は coiled-coil 領域によって規定されるが、その位置はプログラム COILS により容易に予測することができる (Fig.5)。

一方、二量体化促進ドメインには、構造的な特徴がなく、保存性も低いためにその境界を配列情報のみから予測することは出来ない。そこで Gal4 の DNA 結合ドメイン二量体を安定化させて DNA 結合能を向上させることが報告されている  $\lambda$  ファージの cI リプレッサーでこの領域を置換することで、二量体化促進機能を付与することにした。

転写活性化ドメインには酵母細胞内で、翻訳後修飾や他の因子との相互作用等の様々な制御機構が働きかける。そこでそのような制御から全く自由に強力に転写を活性化するドメインとして知られている VP16 でこの領域を置換することにした。(Fig.4)

こうしてデザインされたキメラ転写因子を発現させる為に、プロモーターとしてはガラクトースによって転写を誘導することが可能な GAL1 プロモーターを用いることにした。

これにより、転写因子の量や局在や活性化に必要な相互作用や翻訳後修飾等の上流からのシグナル伝達を考慮する事無く、下流標的遺伝子の誘導を行う事が可能となる (Fig.2)。

### 3.2 キメラ化 Sip4 発現株の作成

今回の実験では、宿主株を容易に変更できるように、発現プラスミド pRS415chimeraZEO-Sip を作成した。

pRS415chimeraZEO-Sip は、GAL1 プロモーターと  $\lambda$ cI と VP16 を持ち ZEO 耐性マーカーを持ったプラスミド pRS415chimeraZEO を元にして、GAL1 プロ

モーターと  $\lambda$ CI リプレッサー遺伝子の間に Sip4 の DNA 結合ドメインをコードする遺伝子断片を挿入することで作成した。まず、酵母ゲノム DNA を鋳型にして、PCR によって Sip4 の DNA 結合ドメインを含む DNA 断片を増幅した。一方で pRS415chimeraZEO を鋳型に、GAL1 プロモーターの下流に Sip4 の開始コドン以下 40bp が連結されたような断片を PCR によって作成した。同様に PCR を用いて、 $\lambda$ CI 遺伝子の上流に Sip4 の coiled-coil 領域をコードする DNA が連結されたような断片を作成した。この 3 種類の断片を制限酵素消化で直線化したベクター pRS415chimeraZEO とともに酵母細胞に形質転換する。そうすると Fig.3 に示すようなギャップ修復反応による組換えが起こり、これらの断片が連結された環状分子（組換えプラスミド）が生じる。このプラスミドを持ったクローンを Zeocin 耐性で選択した。

得られた形質転換体からプラスミドを回収して、塩基配列の決定を行い、変異が入っていない事を確認した。更にこのプラスミドを形質転換した酵母細胞をラフィノース培地で培養後、ガラクトースを添加し GAL1 プロモーターによる発現を誘導した。キメラ転写因子に付加したタグに対する抗体を用いて western blotting を行ったところ、予想されるサイズに一致する約 37kDa のバンドの誘導を確認することが出来た。以上の結果より、キメラ化 Sip4 の発現を誘導できる株が樹立できたと結論した。

### 3-3 キメラ化 Sip4 発現株のマイクロアレイによる発現遺伝子解析

次にキメラ化 Sip4 が細胞内で機能するか否かを明らかにする為に、既知標的遺伝子の発現誘導を検討することにした。Sip4 は試験管内で CSRE 配列に結合することが報告されているので、それを上流に持つ遺伝子 *FBP1* と *SFC1* について検討を行うことにした。ガラクトースの添加によってキメラ化 Sip4 の発現を誘導した後に、RNA を抽出して RT-PCR により *FBP1* と *SFC1* の発現を調べた。その結果、インサートを含まない発現ベクター pRS415chimeraZEO を持つ対照細胞に比べて、両遺伝子ともに有意な mRNA 量の増加が確認された。またガラクトースによる誘導に伴って、標的遺伝子の発現が誘導されることも確認された。以上の結果をより、キメラ化 Sip4 が機能していることが確認されたので、マイクロアレイによる包括的な発現遺伝子解析を行うことにした。

pRS415chimeraZEO を保持した対照細胞と pRS415chimeraZEO-Sip を持つ細胞から抽出した RNA をそれぞれ Cy3 と Cy5 で標識して、出芽酵母の全 ORF がプリントされたマイクロアレイに競合ハイブリダイゼーションを行った。その結果、対照細胞に比してキメラ株で発現が 4 倍以上増強した 42 遺伝子を表 1

にまとめた。これらの遺伝子の中には、*in vitro*において Sip4 が上流に結合すると知られている全ての遺伝子(*FBP1*, *ICL1*, *PCK1*, *SFC1*)が含まれていた。このことからキメラ化 Sip4 が細胞内で機能していると結論付けられた。また、Sip4 との機能の類似性が指摘されている Cat8 の標的と確認されている多くの遺伝子の発現が誘導された。

更にこれら 42 遺伝子の上流配列を解析して、CSRE 配列(CCRTYSRNCCG) および CSRE からそれぞれ 1 塩基、2 塩基ミスマッチが入った上流配列の有無を調べた。この時ミスマッチの位置については特に考慮していない。その結果、42 遺伝子中 20 遺伝子が CSRE 様配列を有している事が見いだされた。しかしながら、発現誘導倍率と CSRE 様配列の存在の間には、単純な相関は見られなかった。

次に、発現上昇が見られた 42 遺伝子に共通する特徴を明らかにする為に、SGD (Saccharomyces Genome Database, <http://www.yeastgenome.org/>) から提供されている GO Term Finder を用いて解析を行った。その結果、Biological Process では糖代謝に関連する GO が有意に濃縮されていることが分かった (Fig.6)。更にこれらの遺伝子を KEGG (<http://www.genome.jp/kegg/>) の代謝マップにプロットした (Fig.7)。これらの結果からキメラ化 Sip4 により糖新生遺伝子が有意に濃縮されていることがわかった。

### 3.4 標的遺伝子濃縮の試み

キメラ化 Sip4 により対照細胞に比べて 4 倍以上発現が誘導された遺伝子の中には CSRE 様配列を持ち直接の標的を持つものが 20 遺伝子、CSRE 様配列を持たず間接の標的あるいはノイズと思われるものが 22 遺伝子含まれていた。キメラ化による標的遺伝子探索戦略を一般化して、機能未知転写因子に適用する為には、ここで得られた 42 遺伝子から CSRE に関する知識を使わずに CSRE 様配列を持つ 20 遺伝子を濃縮できるような方法が必要である。

この問題を解決するために私は下記のようなアルゴリズムを考案し、それに基づくプログラムを作成して解析を行った。(Fig.8)

- 1 キメラ化により発現が 4 倍以上上昇した遺伝子 (標的候補遺伝子) について、パブリックドメインの時系列マイクロアレイデータに基づき k-means 法によるクラスタリングを行う。このときクラスタの数は 3 と設定した。また、ここでは SGD (Saccharomyces Genome Database) から得られる全プロファイル(13 種類)を用いた。
- 2 各プロファイルにおいて、生成されたクラスタ内の遺伝子発現強度の平均値

が最も高いクラスタを“トップクラスタ”とする。(この時点で、13個のトップクラスタが作成される事となる。)

- 3 遺伝子ごとに13個のトップクラスタのうち何個に含まれているか、すなわちトップクラスタに出現する頻度を計算して、それによって順位をつける。

この結果、CSREを上流に持つ8遺伝子の濃縮が認められた。発現量に基づく順位付けではCSREを上流に持つ8遺伝子の順位は42遺伝子中1~22位に分散していたが、トップクラスタ中への出現頻度に基づく順位付けでは8遺伝子が上位9位以内に全てランクされた。その一方で、CSREと1ないし2塩基のミスマッチを持つ遺伝子に関しては、このような濃縮は認められなかった。(表.3)

### 3.5 結合配列解析の試み

前節では遺伝子に主眼を置いていたが、ここでは転写因子が結合するシスエレメントの抽出を試みた。

Zn<sub>2</sub>Cys<sub>6</sub>型転写因子の認識配列は一般にCGGX<sub>n</sub>CCGであることが報告されているので、探し出すコンセンサス配列はトリプレット・{X}<sub>n</sub>・トリプレット(Dyad sequence)に限定することができる。その為の開発されたのが“Dyad”である(5)。しかしこのプログラムは、上流配列を純粹に統計学的に扱うだけのものであるため、これに生物学的な視点からのweightを加えることで精度の向上を図ることにした。

#### 3-5-1 アルゴリズム

##### 3-5-1-1 検定統計値 P-value の算出

キメラ化 Sip4 により対象細胞に比べ4倍以上発現が上昇した42遺伝子の上流に存在する配列のP値を以下に従って算出した。

従来、Zn<sub>2</sub>Cys<sub>6</sub>型転写因子の認識配列はCGGX<sub>n</sub>CCGとされてきたが、研究の進展とともにCSREのように両脇のトリプレットにCとGだけでなくAやTを含む例も知られるようになってきた。しかし、AやTの出現を無制限に許容するとプロモーター領域に存在するTATAボックス等のATリッチな配列ばかりが捉えられてくる。これを避けるために、

- 1 6ベースの内AまたはTは合わせて2文字以内、
- 2 トリプレット中にAAもしくはTTが出現してはならない、

という二つの条件を設定した。

Dyad sequence の P-value の算出は以下に従って行った。

1) トリプレット “w” の出現頻度  $F_{obs}$  の算出には次式を用いた。

$$F_{obs}(w) = Occ(w) / \sum_i [Occ(w_i)]$$

ここで、 $Occ(w)$  はトリプレット  $w$  がクエリー配列中に出現する回数を表す。

又、 $\sum_i [Occ(w_i)]$  はクエリー配列中の全てのトリプレット ( $4^3$  種類) の出現回数の和を表す。

2) Dyad sequence “D” の出現頻度  $F_{exp}$  の算出には次式を用いた。

Dyad sequence “D” は両脇にトリプレット  $w_1, w_2$  を持っているとする

$$F_{exp}(D) = F_{obs}(w_1) \times F_{obs}(w_2)$$

3) Dyad sequence “D” の検定統計値 (P-value) の算出には次式を用いた。

$$P(D, \geq n) = \sum_{i=n}^T {}_T C_i [F_{exp}(D)]^i * [1 - F_{exp}(D)]^{T-i}$$

なお  $T$  は Dyad sequence が出現しうる位置の合計として次式で与えられる

$$T = \sum_j [L_j - K + 1] - Occ(D) * (K - 1)$$

L : クエリーの配列長の合計

K : Dyad sequence の長さ (= スペースの長さ + 3 + 3)

また、Dyad sequence の相補鎖同士は P-value の積をとることによりマージした。

### 3-5-1-2 weight の算出

生物学的知見に基づく weight として以下の 2 つを検討した。

1) 翻訳開始点からの位置情報に基づく重み “Po\_weight”

転写因子の結合位置の傾向を算出して重みとして用いた。Dyad sequence ご

とにそれを持つ遺伝子の集団を G とする。転写因子データベース TRANSFAC (<http://www.gene-regulation.com/pub/databases.html#transfac>) から、出芽酵母の転写因子について翻訳開始点からの結合位置を集計し (Fig.9)、このデータを元に近似曲線  $P_0(x)$  を求めた。

$$Po\_weight = Po(distanceM)$$

distanceM : G における翻訳開始点からの距離の平均  
 $P_0$  : Fig.9 参照

## 2) 遺伝子間の発現類似度に基づく重み “Co\_weight”

1) と同様に Dyad sequence ごとにその配列を持っている全ての遺伝子集団を G とする。G における相関の平均  $G\_corre$  をパブリックドメインに存在する発現プロファイルごとに求める。次に発現の変化量が少ないプロファイルが高い値を持つことを避けるため G における標準偏差の平均  $G(std)$  を算出する。また、各プロファイルの標準偏差の平均  $P(std)$  を算出する。

次式により、プロファイルごとのスコア S を算出し、その最大値を weight とする。(Fig.10)

$$S = G\_corre * (G(std) / P(std))$$

$$Co\_weight = Max \{ S \}$$

最終的に Dyad sequence ごとに与えられる値 “Score” を次式により算出する

$$Score = -\log(P\text{-value}) * Po\_weight * Co\_weight$$

### 3-5-2 Sip4 への適用

キメラ化 Sip4 により対照細胞に比べて 4 倍以上発現が誘導された遺伝子について上記のプログラムを適用した。3-4 と同様に CSRE に着目し本プログラムの有用性を評価した。結果を Fig.11 に示す。この中で、左軸は  $-\log(P\text{-value})$  の値を表し、グラフ中ピンクの折れ線に対応している。また、右軸は 3-5-1 で算出した

“Score”を表し、青の折れ線に対応している。横軸はそれぞれの Dyad sequence をあらわしており、Score 順に並べられている。

その結果、P-value のみを考慮に入れたときに比べて CSRE の濃縮が確認できた (Fig. 11, 黒い矢印)。また P-value の値が低い (つまり  $-\log(\text{P-value})$  の値が高い) が、CSRE ではない配列の値を低下させる効果も見て取れた。(Fig. 11 赤い矢印)

次に、これらをより詳しく解析するために次式により “Index” を定義した。

$$PEAK\_Index = \sum^{10} X / (\sum(X) * RONA)$$

X には Score、 $-\log(\text{P-value})$ 、 $-\log(\text{P-value}) * \text{Po\_weight}$ 、 $-\log(\text{P-val}) * \text{Co\_weight}$  のいずれかの値が代入される。

また、 $\sum^{10}(X)$  は X の上位 10 の和を、 $\sum(X)$  は全ての X の和を、RONA (Ranking Of the Nearest Answer) は X を降順で並び替えた際の最も高い CSRE のランクを、それぞれ表している。なお、上位 10 件は、捕らえられてきた候補配列をレポーターアッセイ等により実験的に検討する際に十分に実現できる値として設定した。

その結果を、Fig. 12 に示す。この結果から、Po\_weight、Co\_weight を加えることにより、P-value に比べて高い PEAK\_Index 値が得られ、生物学的知見に基づく weight を加味することの有用性が確かめられた。

### 3-5-3 他の転写因子への応用 (コンセンサス配列既知の例)

上記プログラムの  $\text{Zn}_2\text{Cys}_6$  型 zinc finger 転写因子一般に対する有効性を評価する目的で、結合配列パターンが既知である 3 種の転写因子 Ppr1、Put3、Ume6 についても同様に解析を行った。

Ppr1 はウラシルの生合成を制御し、標的遺伝子上流に存在する CGGnnWWYKCCG に結合することが知られている。Put3 はプロリン代謝を制御する転写因子であり、CCGnAnGCnAnnnCCG に結合する事が報告されている。Ume6 は初期減数分裂を制御する転写因子であり、TAGCCGCCGA に結合する事が報告されている。

これらの転写因子に関しては、キメラ化による活性化実験の発現データが本研究室で既に得られており、その結果を用いて、上記プログラムによる解析を行った。結果を Fig. 15~17 に示す。どの転写因子においても既知のシスエレメントが高い順位で見出されることが確認できた。



#### 3-5-4 コンセンサス配列未知の転写因子への応用

オレイン酸により活性化することが知られている Oaf1 に本プログラムを適用した(Fig.18)。その結果、非常に高い score で CCC{8}GCA と GCC{7}GGG が検出された。Oaf1 に関しては結合配列が未知であり、この 2 種類の配列は実験的検証の対象となる有力な候補と考えることが出来る。

## 4. 考 察

本研究において、私はキメラ化 Sip4 発現株を作成し、マイクロアレイ解析を行い、現在まで *in vivo* において標的の解明が進んでいなかった Sip4 転写因子についてその標的候補遺伝子の同定を試みた。

Sip4 は Zn<sub>2</sub>Cys<sub>6</sub> 型の転写因子であり、*in vitro* で Carbon-Source Response Element (CSRE) に結合することが知られており、糖新生に関与する *FBP1*, *ICL1*, *PCK1*, *SFC1* などの発現を制御することが知られていた。しかし、その標的遺伝子を網羅的に探索した例はなかった。

キメラ化 Sip4 を用いて同定された標的候補遺伝子は、これまでに既知の標的と考えられていた *FBP1*, *ICL1*, *PCK1*, *SFC1*) を含んでいた。一方、グルコース存在下で Sip4 が Snf1 セリンスレオニンプロテインキナーゼにより活性化されて、K<sup>+</sup> トランスポーター (Trk1, Trk2) の発現を制御するという報告がある (4) が、これらの遺伝子の有意な発現上昇は見られなかった。これらの遺伝子上流に CSRE を持っていないことも考慮するとこれらは直接的な標的ではなく、二次的な標的である可能性も強い

キメラ化 Sip4 を用いて同定された標的候補遺伝子に対して、GO Term finder での解析及び KEGG の代謝マップへのプロットを行ったところ、これらの多くが糖代謝関連遺伝子であることが分かった。そして、これらの糖代謝関連遺伝子の多くが上流に CSRE 様配列を持っている事が見いだされた。これらの結果は、これらが Sip4 の直接の標的遺伝子であることを示唆しており、Sip4 の機能が糖代謝の調節であることを更に裏付ける結果であると考えられた。

近年、転写因子の標的遺伝子探索の有効な手法としてクロマチン免疫沈降法が注目されている。Sip4 に関しても Young らによる大規模 ChIP-Chip 解析の一環として解析がなされている。しかしながら、その結果からは、糖代謝に関係した標的遺伝子は捕捉されていない。これは培養条件が Sip4 の標的遺伝子への結合に重要な影響を与える為であり、そこを無視して不適切な条件下で ChIP を行っても意味がないことを示唆している。したがって、適切な活性化条件を先見的に知りえない転写因子に関しては、ChIP よりもキメラ化による活性化の方が遥かに有効であることを物語っている。

Sip4 と相同性が高い転写因子として Cat8 が知られており、非発酵性栄養条件化で機能し、標的遺伝子の CSRE に結合すると知られている。Sip4 と Cat8 の機能分化について知見を得ることを目的として、Cat8 に関しても Sip4 と同様の戦略でキメラ化転写因子発現株の作成を行った。

しかしながらキメラ化 Cat8 発現株では RT-PCR、マイクロアレイ解析のいずれにおいても既知標的の発現の上昇を確認する事ができなかった。原因として転写因子が核に移行していない可能性が考えられた。そこでキメラ化 Cat8 に更に NLS (nuclear localization signal) を付加した株を作成し、RT-PCR を行ったが、既知標的の発現上昇は見られなかった。この原因は特定出来ていないが、キメラ化 Cat8 のタンパク質の高次構造に問題が生じ、DNA に結合できないもしくは転写の活性化ができないコンフォメーションをとっている可能性も考えられる。一次配列上で相同性が高い Sip4 と Cat8 の間で、キメラ化たんぱく質の作成においてこのような差異が生じることは、キメラ化による戦略を進める上で留意しなくてはならない点であろう。

Sip4 の直接の標的遺伝子は、上流に CSRE 様配列を持っていると考えられるが、表.1 で見取れるように標的候補遺伝子中には CSRE 様配列を持っていない遺伝子も多数含まれる。これらの CSRE を持たない遺伝子は Sip の直接の標的ではなく二次的な標的、つまり Sip4 により発現が誘導される遺伝子により転写が活性化される遺伝子群もしくは更にその下流遺伝子であると考えられる。この他、Sip4 の標的ではなく、非特異的に発現が誘導された遺伝子もこの中に含まれると考えられる。これらの遺伝子は Sip4 存在下で発現が上昇した遺伝子群についての Go Term Finder での解析や、上流配列解析を行う際のノイズとなる。そこで直接の標的遺伝子と副次的なものとを識別する必要がある。

ひとつの手段は、キメラ転写因子に対して ChIP-Chip を行う実験的なアプローチであり、もう一方のアプローチはバイオインフォマティクスのなものである。後者に関しては、CSRE 様配列の有無と誘導倍率の間に単純な相関はないので、この実験データだけから直接の標的遺伝子を抽出することは不可能なである。そこで発現量以外の情報を利用した Sip4 の直接標的の抽出を試みた。

まず、パブリックドメイン上のマイクロアレイデータを用いクラスタリングを行い、Sip4 の標的遺伝子の濃縮を試みた。これは「同一転写因子により発現が制御される遺伝子群は様々な環境化で同じ挙動を示す」という仮定に基づいている。結果として CSRE 配列を持つ遺伝子が有意に濃縮される結果となった。これは先の仮定を支持するものと考えられる。

次に統計学的に上流配列を解析するアルゴリズムに、生物学的知見を weight として加味したプログラムの作成を行った。

第一の weight としては、翻訳開始点までの位置情報を取り入れた。これは転写因子がシスエレメントに結合し、RNA ポリメラーゼと相互作用することにより転写活性を制御するというメカニズムを考慮すると、出芽酵母の転写因子の結合可能な領域が限定的であると考えたからである。(本来は、転写開始点情報

を利用すべきであるが、酵母ではそのデータが大きく欠落しているため、今回は ORF の開始コドンを利用した。また、哺乳類細胞では転写開始点上流数十 kb から転写を制御するエンハンサーが知られているので、本プログラムを適用する際には翻訳開始点からの位置情報を再考する必要があると考えられる。）

現在、翻訳開始点からの距離による **weight (Po\_weight)** は同一エレメントを持つ遺伝子集団 **G** における平均値を用いている。**Index** を導入して **Score** への寄与を評価したところ、**Po\_weight** の寄与は非常に小さいことが確認された。これの改善にあたっては、距離による **weight** を各遺伝子間の相関を求める際に付加することでノイズを軽減できる可能性がある (**Appendix** 参照)。また翻訳開始点ではなくて、転写開始点からの位置情報を用いることにより、位置情報による **Score** への寄与を増大させ、シスエレメントの予測精度の向上につながる可能性がある。

第二の **weight** として遺伝子発現パターンの類似度 (**Co\_weight**) を用いた。現在は、SGD から取得できるプロファイルのみを使用しているが、発現プロファイルの種類が増加するほど、シスエレメントの予測精度の向上が期待できる。なぜならば対象となる転写因子が活性化される条件に近いプロファイルが存在すれば、それによって高い相関係数が得られるからである。現在、種々の条件におけるマイクロアレイデータが急速に蓄積されておりこれらのデータを用いることを検討している。

プロファイル数の増加によりノイズの影響が懸念されるが、アルゴリズムの構造上、具体的には遺伝子間の相関が最も高い値を示すプロファイルのみを使用するため、数の増加による影響に対してはロバストであると考えられる。また、各 **Dyad sequence** を持つ遺伝子集団 **G** において最も高い相関を示すプロファイルより、**G** がどのような条件で発現が誘導されるかを知る手がかりになるかもしれない。

これら 2 つの生物学的知見に基づく **weight** を取り入れた **Score** を用いることで、**Dyad** 原法と同様に **P-value** だけに依存していた場合よりも、標的配列の抽出精度を向上できることが、**Sip4** のみならず他の転写因子に関しても示された。

さて、上記のプログラムによって、標的配列が上位にランクされることは示された。しかし、実験的検証を行うには、上位集団から更に候補の絞込みが出来ることが望ましい。そこで各 **Dyad sequence** をプロモータ領域を持つ遺伝子群の機能の類似度を考慮に入れることを検討している。もしその **Dyad sequence** が真のエレメントであれば、それを持つ遺伝子群の機能が相互に類似すると期待されるからである。このアプローチの有効性を検証する予備実験として、各

Dyad sequence を上流に持つ遺伝子群を GO Term Finder で解析し、biological process において計算された最も低い Probability を重みとして加味することを検討した。具体的には、次式を用いて重み BP\_weight を算出した。

$$BP\_weight = -\log(\min\{Probability\})$$

この式に基づいて Sip4 の解析で上位 10 位にランクされた Dyad sequence について BP\_weight を算出し、それを Score に乗じたもので評価を行った (Fig.13)。(各 BP\_weight は Fig.14 に示す。) BP\_weight を加えることによって、Score だけにに基づく場合 (Fig.11) よりも、CSRE がより上位にランクされることが分かったので、Score の算出自体に BP\_weight を組み込むことも検討する価値があると思われた。

また、上位にランクされた Dyad sequence の中から真の標的配列を見出す際に、比較ゲノムデータを活用することも検討している。真の標的配列で機能的なものであるならば、近縁種での保存が期待できるので、オルソログにおける Dyad sequence の保存を評価に利用できるであろう。出芽酵母に関しては、現在近縁 6 種 (*S. paradoxus*, *S. mikatae*, *S. bayanus*, *S. kudriavzevii*, *S. castelii*, *S. kluyveri*) の配列が利用可能であるので、これも今後の検討が可能であろう。

これまで、上位にランクされた Dyad sequence から真の標的を選択することを考えてきた。しかし、そもそも真の標的配列以外に高い Score で現われてきた配列とは何ものなのだろうか？解析対象の転写因子 A が別の転写因子 X を誘導した結果として二次的な標的が検出されている場合には、A ではなくて X の認識配列が浮上してくる可能性も否定できない。また、転写因子の中には標的遺伝子をシェアしている、つまりペアで働く例も少なからず知られている。そのような場合には、解析対象ではない方の転写因子の標的配列が浮上している可能性も高い。これらのケースを考慮すると、上位にランクされた Dyad sequence に関しては注意深い観察が必要であろう。

一方、キメラ化戦略は人工的蛋白質を細胞内に過剰に発現させるものなので、そのこと自体によるストレスを宿主細胞が感じて応答していることも忘れてはなるまい。キメラ転写因子を発現させた際のプロファイルが蓄積してくれば、自ずと共通に誘導される遺伝子群が浮上してくるので、それをマスクすることで解決が可能であると考えられる。既に本研究室では全 Zn<sub>2</sub>Cys<sub>6</sub> 転写因子のキメラ化実験データが出揃いつつあるので、この問題にも取り組んでみたい。

最後に、現在のところ、上位にランクされた **Dyad sequence** はそれぞれ別のものと見做して取り扱っているが、それらは或るコンセンサスとしてまとめられるべきものかも知れない。そのようなコンセンサス配列が見えるようになると、データがすっきりと整理されて解釈できるようになるケースも少なくない筈である。しかし、塩基の縮重を無制限に許すと、どんなものでもコンセンサスとして括れてしまう。ここで重要になるのは、ヒトの目ではなくて、転写因子の目にそれぞれの配列がどのように映っているのか、を理解することであろう。対象とする転写因子ファミリーの分子認識機構に関する構造生物学的知見も盛り込んで、コンセンサス配列の導出に取り組むことが今後は必要なのではないだろうか。

これらの残された課題を解決し、改良を加えてゆくことによって、キメラ化による標的遺伝子探索法をより有効な戦略へと進化させ、遺伝子ネットワークの解明に貢献することが出来ると思われる。

## Appendix

Dyad sequence が遺伝子 A,B,C の上流に含まれていると仮定する

$$\text{CoPo\_weight} = \text{Po}(A) \text{Po}(B) * (\text{Correlation\_AB}) + \text{Po}(A) \text{Po}(C) * (\text{Correlation\_AC}) \\ + \text{Po}(B) \text{Po}(C) * (\text{Correlation\_BC})$$

$$* \text{Correlation\_XY} = \max \{G\_corre\}$$

G\_corre: 遺伝子集団 G(この場合 X,Y)の各プロファイルにおける相関係数

## 5. 参考文献

- (1) Onda, M., Ota, K., Chiba, T., Sakaki, Y. and Ito, T.  
Analysis of gene network regulating yeast multidrug resistance by artificial activation of transcription factors: involving of Pdr3 in salt tolerance  
*Gene* **332**: 51-59 (2004)
- (2) Vincent, O. and Carlson, M.  
Sip4, a Snf1 kinase-dependent transcriptional activator, binds to the carbon source-responsive element of gluconeogenic genes  
*EMBO J.* **17**: 7002-7008 (1998)
- (3) Portillo, F., Ulet, J. M. M. and Serrano, R.  
A role for the non-phosphorylated form of yeast Snf1: tolerance to toxic cations and activation of potassium transport  
*FEBS Lett.* **579**: 512-516 (2005)
- (4) Roth, S. and Schuller, H.-J.  
Cat8 and Sip4 regulate transcriptional activation of the yeast malate dehydrogenase gene *MDH2* by three carbon source-responsive promoter elements  
*Yeast* **18**: 151-162 (2001)
- (5) van Helden, J., Rios, A. F. and Collado-Vides, J.  
Discovering regulatory elements in non-coding sequences by analysis of spaced dyads  
*Nucleic Acids Res.* **28**: 1808-1818 (2000)
- (6) Haurie, V., Perrot, M., Mini, T., Jenö, P. and Sagliocco, F.  
The Transcriptional Activator Cat8p Provides a Major Contribution to the Reprogramming of Carbon Metabolism during the Diauxic Shift in *Saccharomyces cerevisiae*  
*J. Biol. Chem.* **276**: 76-85 (2001).
- (7) Sasaki, H. and Uemura, H.  
Influence of low glycolytic activities in *gcr1* and *gcr2* mutants on the expression of other metabolic pathway genes in *Saccharomyces cerevisiae*

*Yeast* **22**:111-127 (2005)

- (8) Elton T. Young, Kenneth M. Dombek, Chris Tachibana, and Trey Ideker  
Multiple Pathways Are Co-regulated by the Protein Kinase Snf1 and the  
Transcription Factors Adr1 and Cat8  
*J. Biol. Chem.* **278**: 26146-26158 (2003)
- (9) COILS-Prediction of Coiled Coil Regions in Protein  
<http://www.ch.embnet.org/software/COILSform.html>
- (10) Richard J. R. and Mark P.  
Determinants of Binding-Site Specificity Among Yeast C<sub>6</sub> Zinc Cluster Proteins  
*Science* **261**: 909-911 (1993)
- (11) Schjerling, P. and Holmberg, S.  
Comparative amino acid sequence analysis of the C<sub>6</sub> zinc cluster family of  
transcriptional regulators  
*Nucleic Acids Res.* **24**: 4599-4607 (1996)



## 6. 謝辞

本研究を進めるにあたり、私の研究を支えてくださった下記の皆様に厚く御礼申し上げます。

指導教官の伊藤隆司先生には、大変お世話になりました。また、中間発表、修士論文提出において提出が遅く申し訳ありませんでした。実験の指導からプログラムのアイデア、修士論文のチェック、その他多岐に渡り太田一寿助手には大変お世話になりました。ありがとうございました。恩田美雪さんには大学院入学当初から実験の指導をはじめ、プレゼン、論文作成にいたるまでご指導いただき大変感謝しています。ありがとうございました。D2の岡田悟さんには、授業の課題にはじまり、プログラミング、修士論文にいたるまで大変お世話になり、ありがとうございました。紀藤圭治助手、三浦史仁さん、川口典子さんにはセミナー等で貴重なアドバイスをいただきました。ありがとうございました。久保田浩行助手、山口佳洋さん、ひょうさん、千葉智子さんには実験、その他様々なことについて貴重なご意見をいただきました。大変感謝しております。瀧沢聖子さんには研究室の運営に関してお世話になりました。ありがとうございました。また、卒業された、松尾さん、史さんにも大変お世話になりました。ありがとうございました。

研究室以外でも、いろいろな方にお世話になりました。中谷明弘先生にはお忙しい中プログラムに関する貴重なアドバイスをいただき、ありがとうございました。また、有田正規先生には修士論文の提出に際してお忙しい中貴重なご意見を頂きありがとうございました。

最後に、現在まで学業を支えていただき、気随な私を許してくれた両親に感謝します。ありがとうございました。

## 7. 発表状況

学会発表

谷 祐治, 恩田美雪、太田一寿、伊藤隆司

**“Analysis of yeast gene regulatory networks: Target gene search of CAT8 and SIP4 by chimerization-mediated activation of transcription factors and expression profiling”**

第28回 日本分子生物学会年会

## 8. 表と図

- Fig.1 遺伝子制御ネットワーク
- Fig.2 キメラ化転写因子の概要
- Fig.3 ギャップ修飾反応によるコンストラクトの作成
- Fig.4 Zn2Cys6 型転写因子のドメイン構造及びキメラ化転写因子のドメイン構造
- Fig.5 COIL による coiled-coil 領域の予測
- Fig.6 Go Term Finder による解析
- Fig.7 KEGG の代謝マップへのプロット
- Fig.8 クラスタリングを用いた標的遺伝子濃縮 (アルゴリズム)
- Fig.9 翻訳開始点からの距離と転写因子の結合頻度
- Fig.10 遺伝子間の発現類似度に基づく重み “Co\_weightt”
- Fig.11 Sip4 における Score 及び P-val
- Fig.12 重みの効果
- Fig.13 BP\_weightt の効果
- Fig.14 各 BP\_weight
- Fig.15 Ppr1 へのプログラム適用
- Fig.16 Put3 へのプログラム適用
- Fig.17 Ume6 へのプログラム適用
- Fig.18 Oaf1 へのプログラム適用
- 表 1 本実験での使用プライマー
- 表 2 キメラ化 Sip4 により発現が誘導された遺伝子
- 表 3 クラスタリングを用いた標的遺伝子濃縮 (結果)

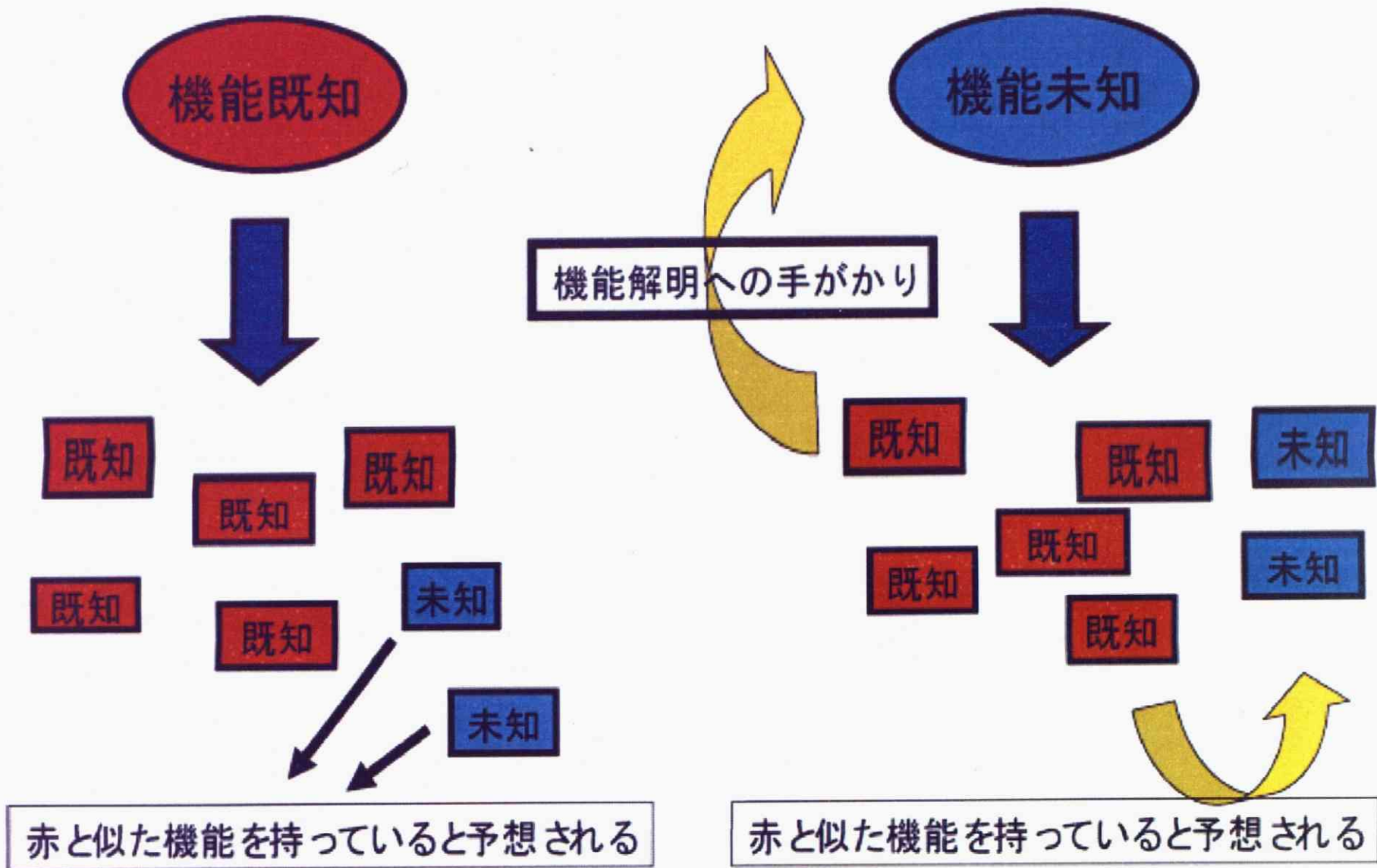


Fig. 1 遺伝子制御ネットワーク

機能未知の遺伝子が機能既知の転写因子により制御されている事がわかればその機能を推測することができる (左図)  
 機能未知の転写因子は標的遺伝子の機能から推測することができる (右図)

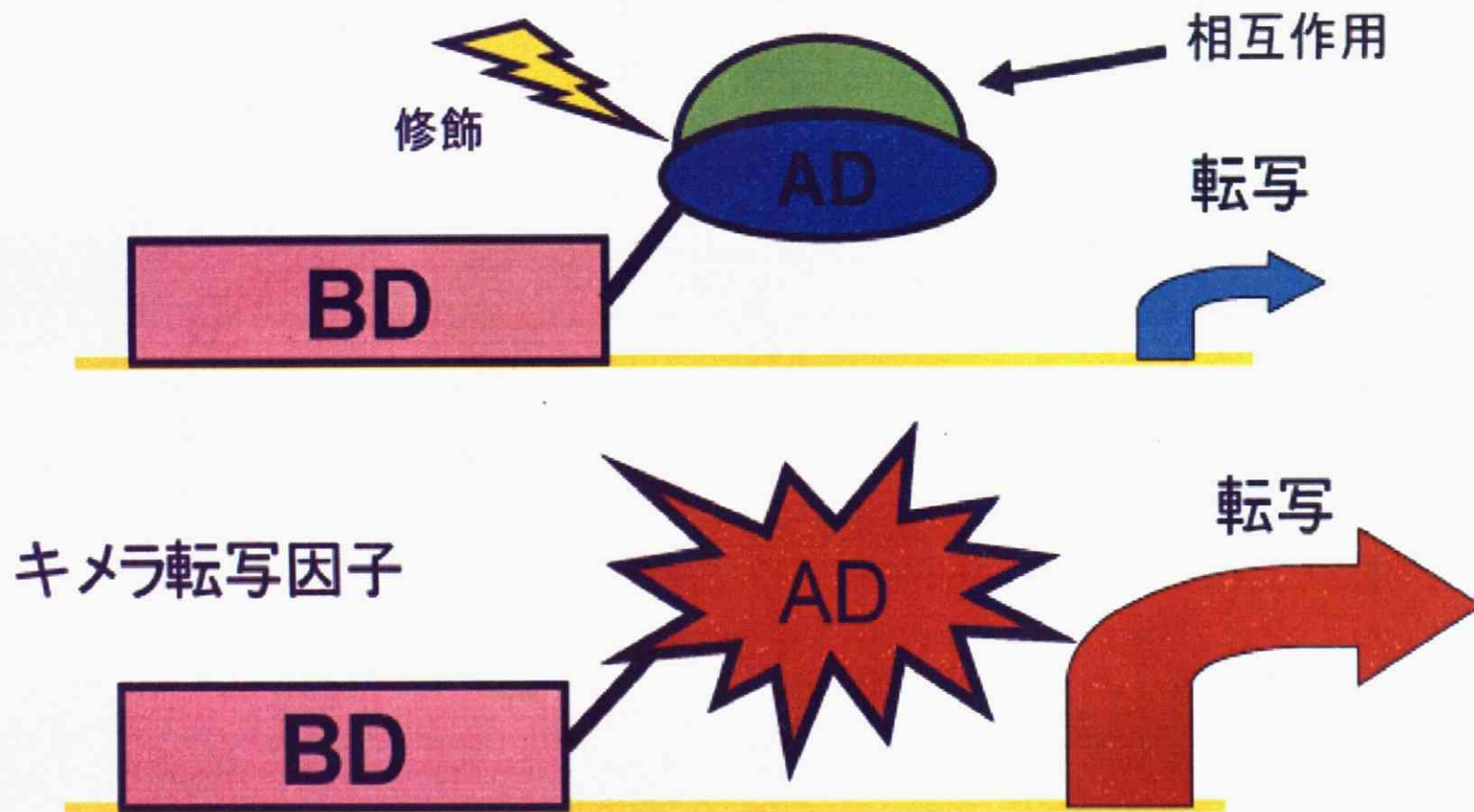
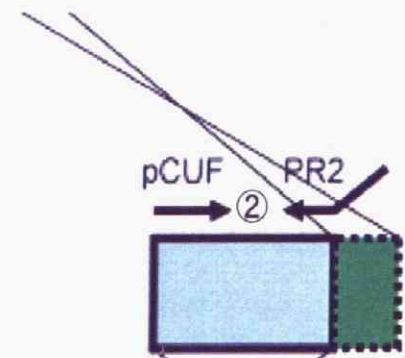


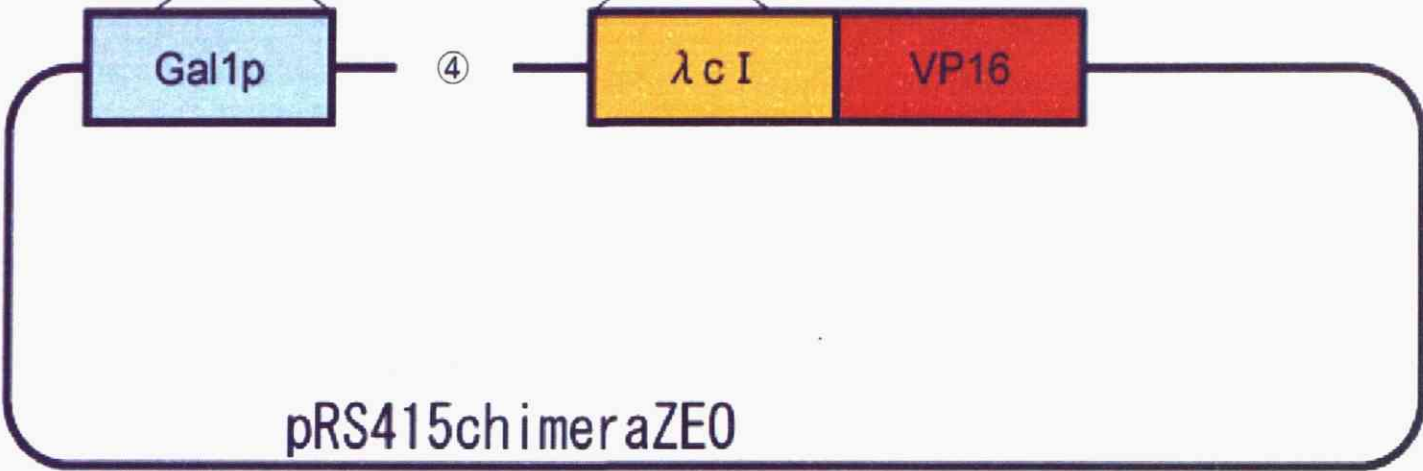
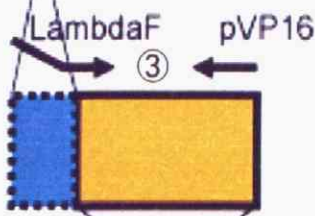
Fig. 2 キメラ化転写因子の概要

恒常的な転写活性を持ち、上流からの刺激などを考慮すること無く標的遺伝子の転写量を増大することができる

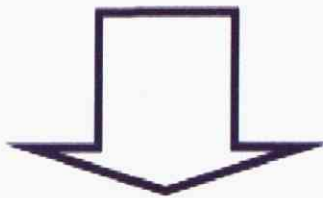
CF1  
①



CF2  
①



pRS415chimeraZE0





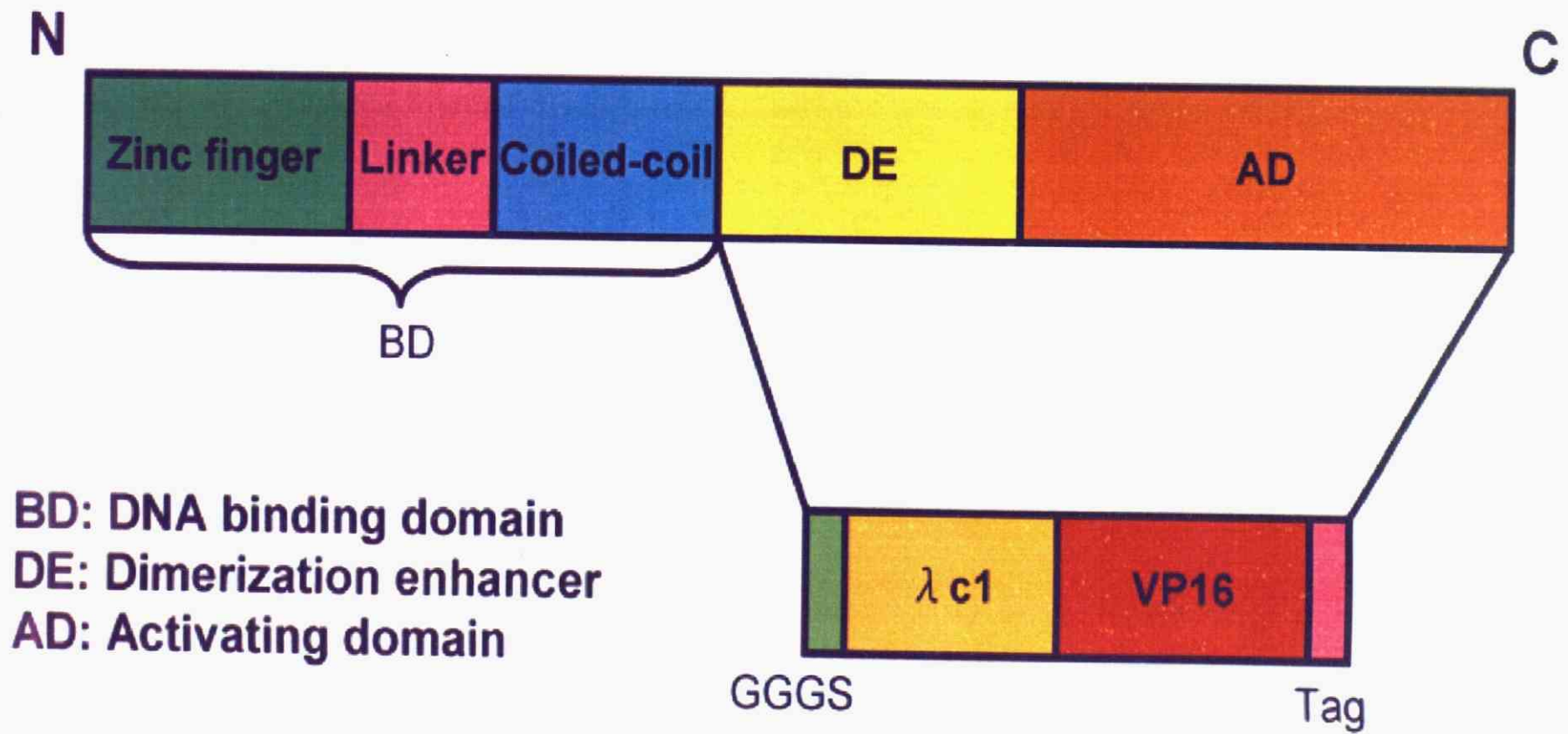


Fig. 4 Zn<sub>2</sub>Cys<sub>6</sub>型転写因子のドメイン構造 (上) 及びキメラ化転写因子のドメイン構造 (下)

Coils output for unknown

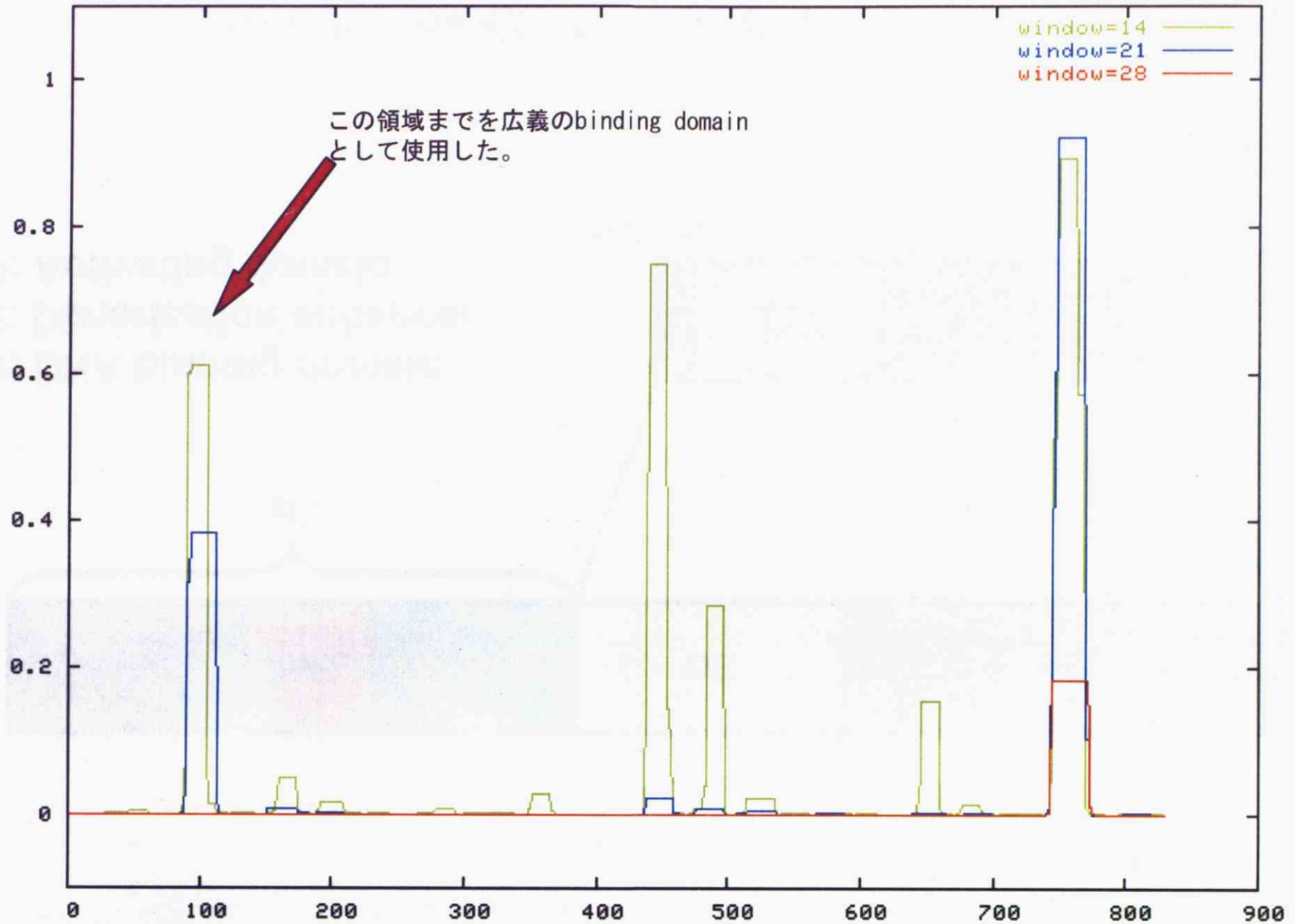


Fig. 5 COILによるcoiled-coil領域の予測  
値が大きいほどcoiled-coil領域である可能性が高い  
横軸はSip4のタンパクの全長を表している



Terms from the Process Ontology

Gene Ontology term	Cluster frequency	Genome frequency of use	P-value	Genes annotated to the term
<a href="#">main pathways of carbohydrate metabolism</a>   <a href="#">AmiGO</a>	8 out of 40 genes, 20%	70 out of 7274 annotated genes, 0.9%	4.29e-09	<a href="#">PCK1</a> , <a href="#">ICL1</a> , <a href="#">FBP1</a> , <a href="#">MDH2</a> , <a href="#">MLS1</a> , <a href="#">CIT2</a> , <a href="#">RKI1</a> , <a href="#">UBC8</a>
<a href="#">alcohol metabolism</a>   <a href="#">AmiGO</a>	9 out of 40 genes, 22.5%	157 out of 7274 annotated genes, 2.1%	1.51e-07	<a href="#">PCK1</a> , <a href="#">YAT2</a> , <a href="#">YAT1</a> , <a href="#">FBP1</a> , <a href="#">MDH2</a> , <a href="#">ERG13</a> , <a href="#">RKI1</a> , <a href="#">UBC8</a> , <a href="#">YPR1</a>
<a href="#">carnitine metabolism</a>   <a href="#">AmiGO</a>	3 out of 40 genes, 7.5%	3 out of 7274 annotated genes, 0.0%	6.85e-07	<a href="#">YAT2</a> , <a href="#">CAT2</a> , <a href="#">YAT1</a>
<a href="#">cellular carbohydrate metabolism</a>   <a href="#">AmiGO</a>	9 out of 40 genes, 22.5%	201 out of 7274 annotated genes, 2.7%	1.18e-06	<a href="#">PCK1</a> , <a href="#">ICL1</a> , <a href="#">FBP1</a> , <a href="#">MDH2</a> , <a href="#">MLS1</a> , <a href="#">CIT2</a> , <a href="#">RKI1</a> , <a href="#">UBC8</a> , <a href="#">YPR1</a>
<a href="#">glyoxylate metabolism</a>   <a href="#">AmiGO</a>	3 out of 40 genes, 7.5%	4 out of 7274 annotated genes, 0.0%	1.61e-06	<a href="#">ICL1</a> , <a href="#">MLS1</a> , <a href="#">CIT2</a>
<a href="#">glyoxylate cycle</a>   <a href="#">AmiGO</a>	3 out of 40 genes, 7.5%	4 out of 7274 annotated genes, 0.0%	1.61e-06	<a href="#">ICL1</a> , <a href="#">MLS1</a> , <a href="#">CIT2</a>
<a href="#">carbohydrate metabolism</a>   <a href="#">AmiGO</a>	9 out of 40 genes, 22.5%	219 out of 7274 annotated genes, 3.0%	2.38e-06	<a href="#">PCK1</a> , <a href="#">ICL1</a> , <a href="#">FBP1</a> , <a href="#">MDH2</a> , <a href="#">MLS1</a> , <a href="#">CIT2</a> , <a href="#">RKI1</a> , <a href="#">UBC8</a> , <a href="#">YPR1</a>
<a href="#">carboxylic acid metabolism</a>   <a href="#">AmiGO</a>	10 out of 40 genes, 25%	300 out of 7274 annotated genes, 4.1%	3.85e-06	<a href="#">PCK1</a> , <a href="#">ICL1</a> , <a href="#">YAT2</a> , <a href="#">CAT2</a> , <a href="#">YAT1</a> , <a href="#">FBP1</a> , <a href="#">MDH2</a> , <a href="#">MLS1</a> , <a href="#">CIT2</a> , <a href="#">UBC8</a>
<a href="#">organic acid metabolism</a>   <a href="#">AmiGO</a>	10 out of 40 genes, 25%	300 out of 7274 annotated genes, 4.1%	3.85e-06	<a href="#">PCK1</a> , <a href="#">ICL1</a> , <a href="#">YAT2</a> , <a href="#">CAT2</a> , <a href="#">YAT1</a> , <a href="#">FBP1</a> , <a href="#">MDH2</a> , <a href="#">MLS1</a> , <a href="#">CIT2</a> , <a href="#">UBC8</a>
<a href="#">energy derivation by oxidation of organic compounds</a>   <a href="#">AmiGO</a>	8 out of 40 genes, 20%	195 out of 7274 annotated genes, 2.6%	9.51e-06	<a href="#">PCK1</a> , <a href="#">ICL1</a> , <a href="#">FBP1</a> , <a href="#">MDH2</a> , <a href="#">MLS1</a> , <a href="#">CIT2</a> , <a href="#">RKI1</a> , <a href="#">UBC8</a>

<http://db.yeastgenome.org/cgi-bin/GO/goTermFinder>

Fig. 6 Go Term Finder による解析

biological process における結果を示している。この結果 Metabolism, 特に糖代謝関連遺伝子の発現がトビ出ていることが目いされた。



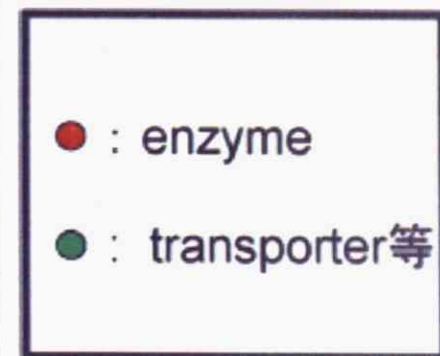
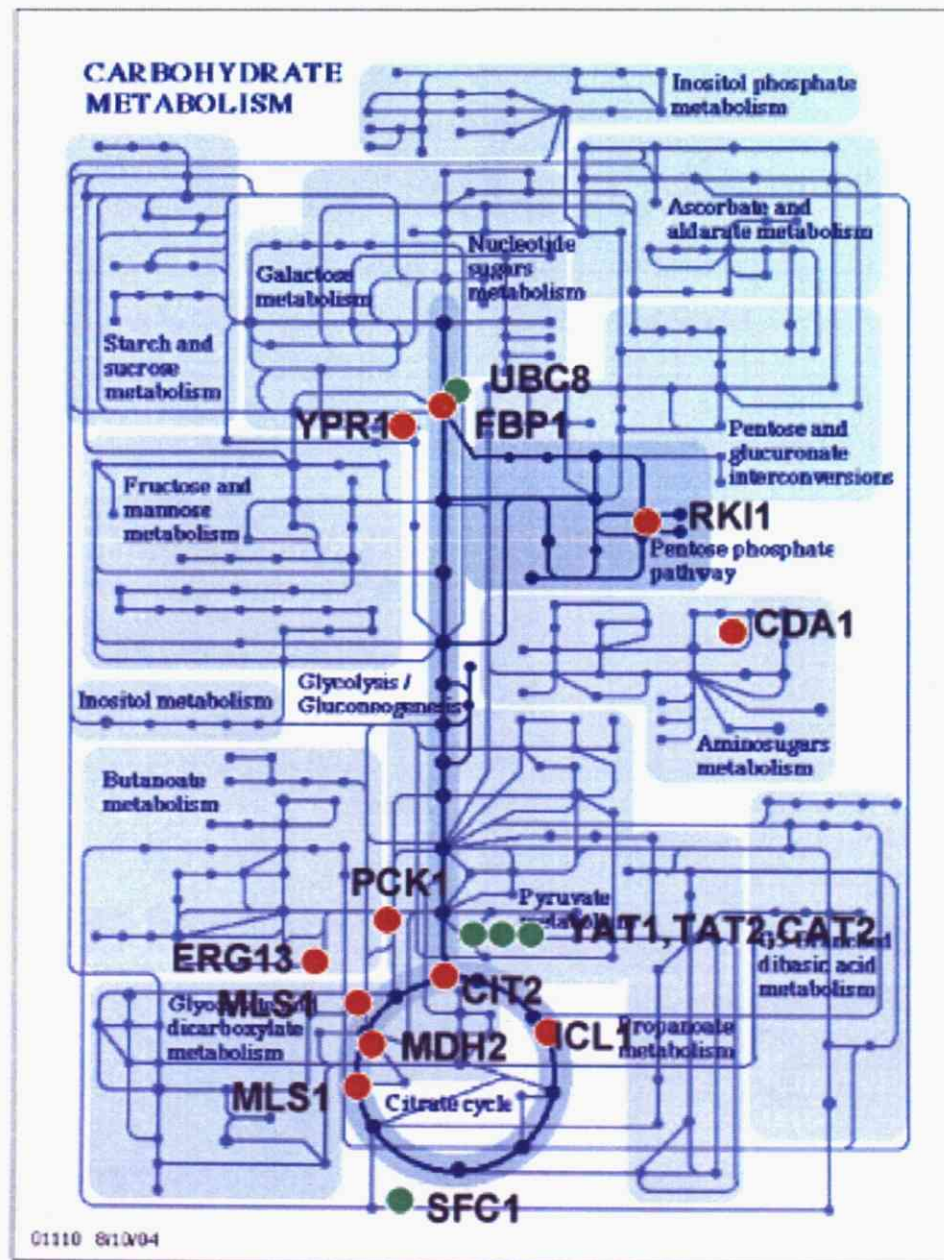


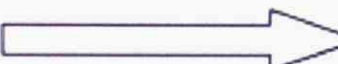
Fig. 7 KEGGの代謝マップへのプロット  
 キメラ化Sip4により発現が上昇する遺伝子の多くが  
 の糖代謝関連遺伝子であった

発現が上昇  
 した遺伝子群

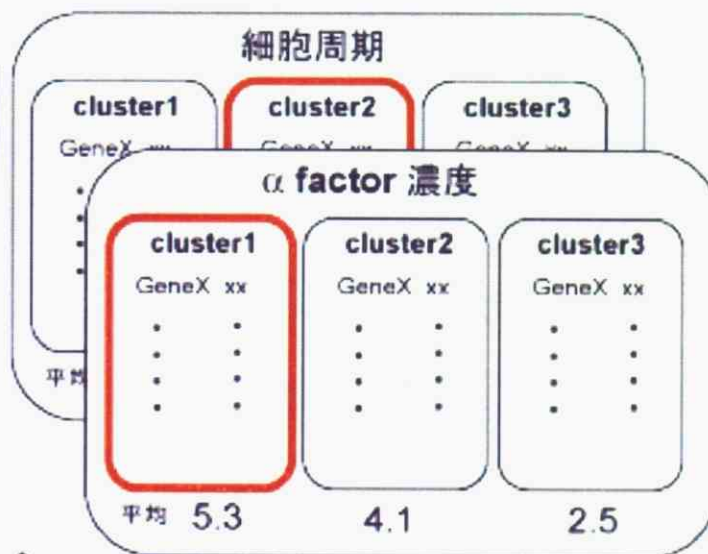
発現が上昇した遺伝子群

Gene1	11.2
Gene2	8.3
Gene3	7.5
Gene3	7.2
Gene4	6.5
Gene5	5.5
⋮	⋮
⋮	⋮
⋮	⋮

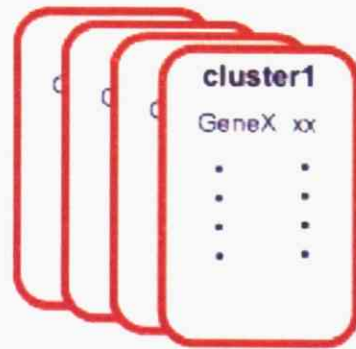
様々な発現データを元にクラスタリング



α factor 濃度  
細胞周期  
リン酸濃度  
DNA 損傷  
⋮  
⋮



平均発現強度が最も高いクラスター (トップクラスター)



発現データごとのトップクラスター

各遺伝子について出現回数をカウントする

geneX	出現回数
⋮	⋮
⋮	⋮
⋮	⋮
⋮	⋮
⋮	⋮
⋮	⋮
⋮	⋮

トップクラスター回数

Fig. 8 クラスタリングを用いた標的遺伝子濃縮

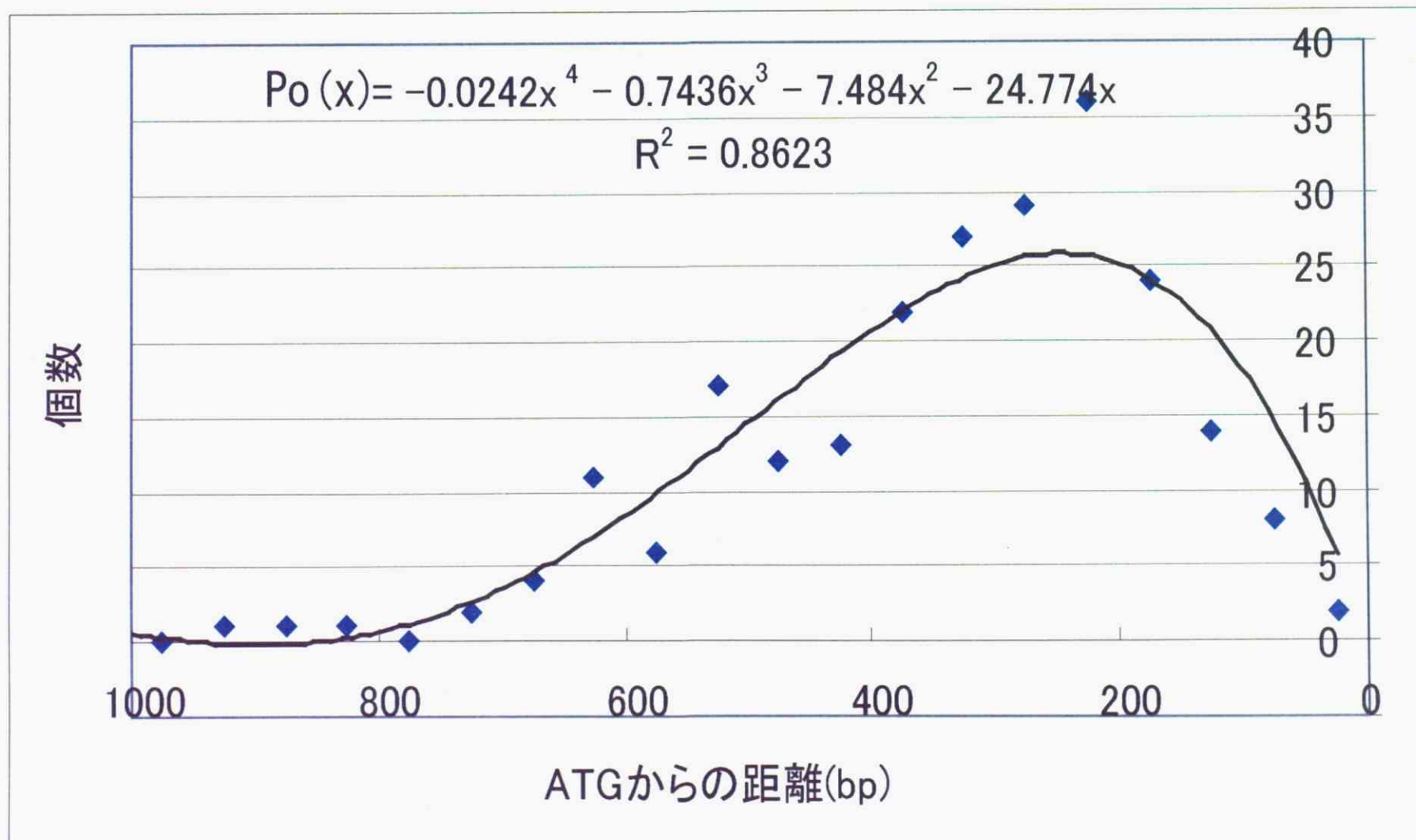


Fig. 9 翻訳開始点からの距離と転写因子の結合頻度

Transfac における *S. cerevisiae* のデータを用いて近似曲線を算出した

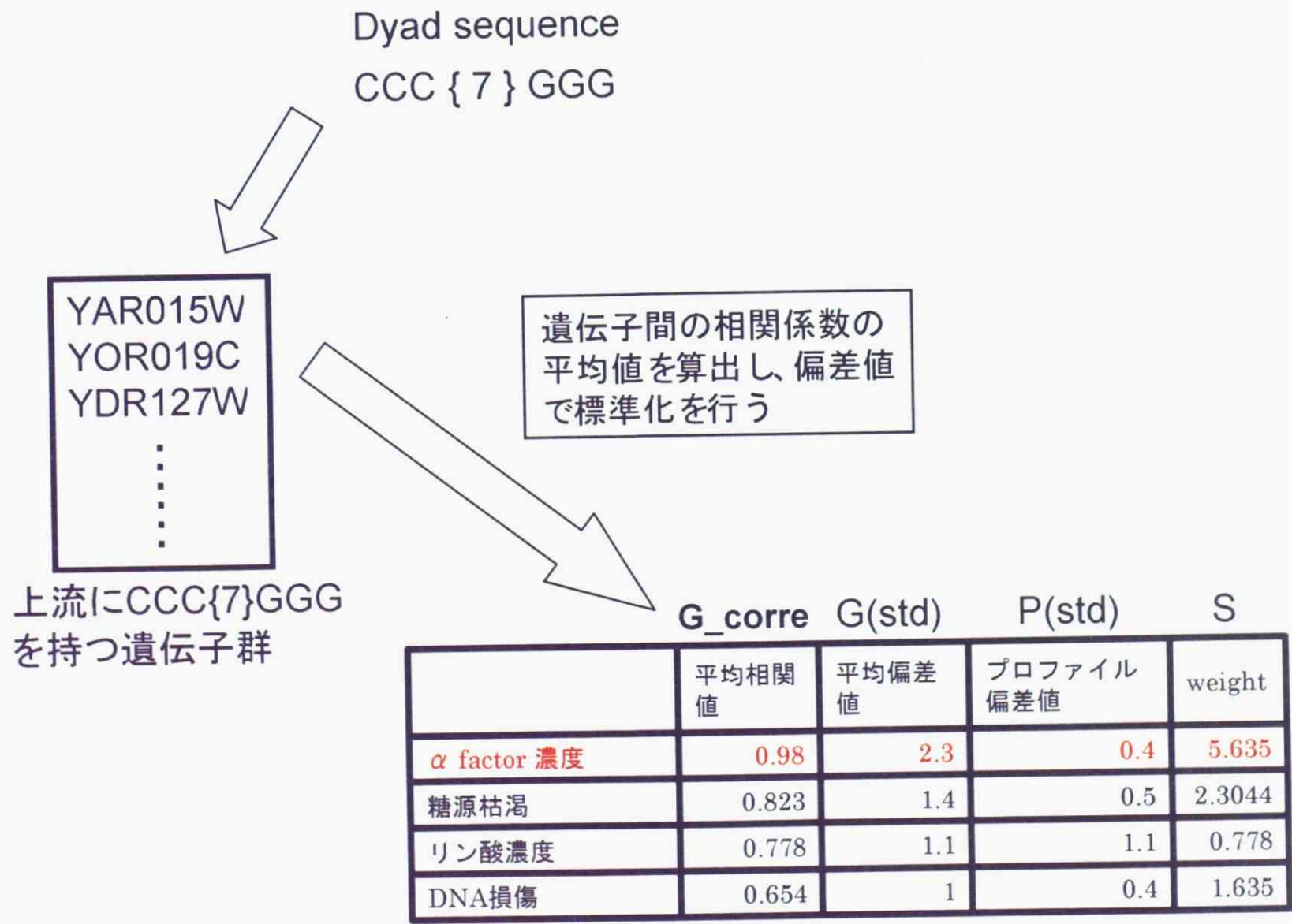


Fig. 10 遺伝子の発現類似度に基づく重み “ Co\_weight ”



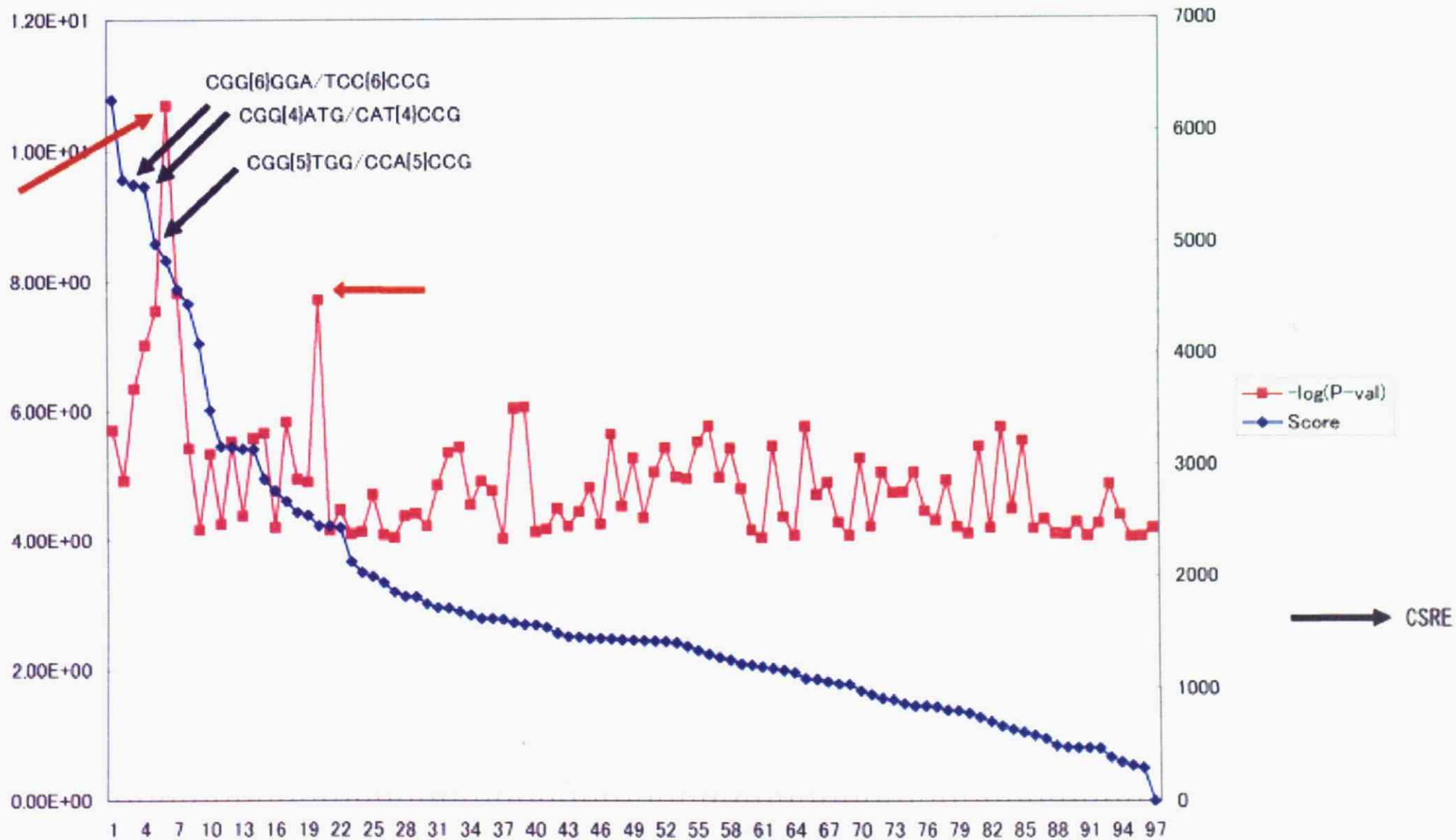


Fig. 11 Sip4における Score および P-val

左軸は $-\log(P\text{-val})$ を表し、ピンクのグラフに対応する

右軸はScoreを表し、青のグラフに対応する

X軸はそれぞれDyad sequenceを表している。

黒い矢印は、Sip4が結合することが報告されているCSREである

赤い矢印で示した様に、低いP-valを持つ非認識配列のScoreが

赤い矢印で示した様に、低いP-valを持つ非認識配列のScoreが抑えられていることが確認された

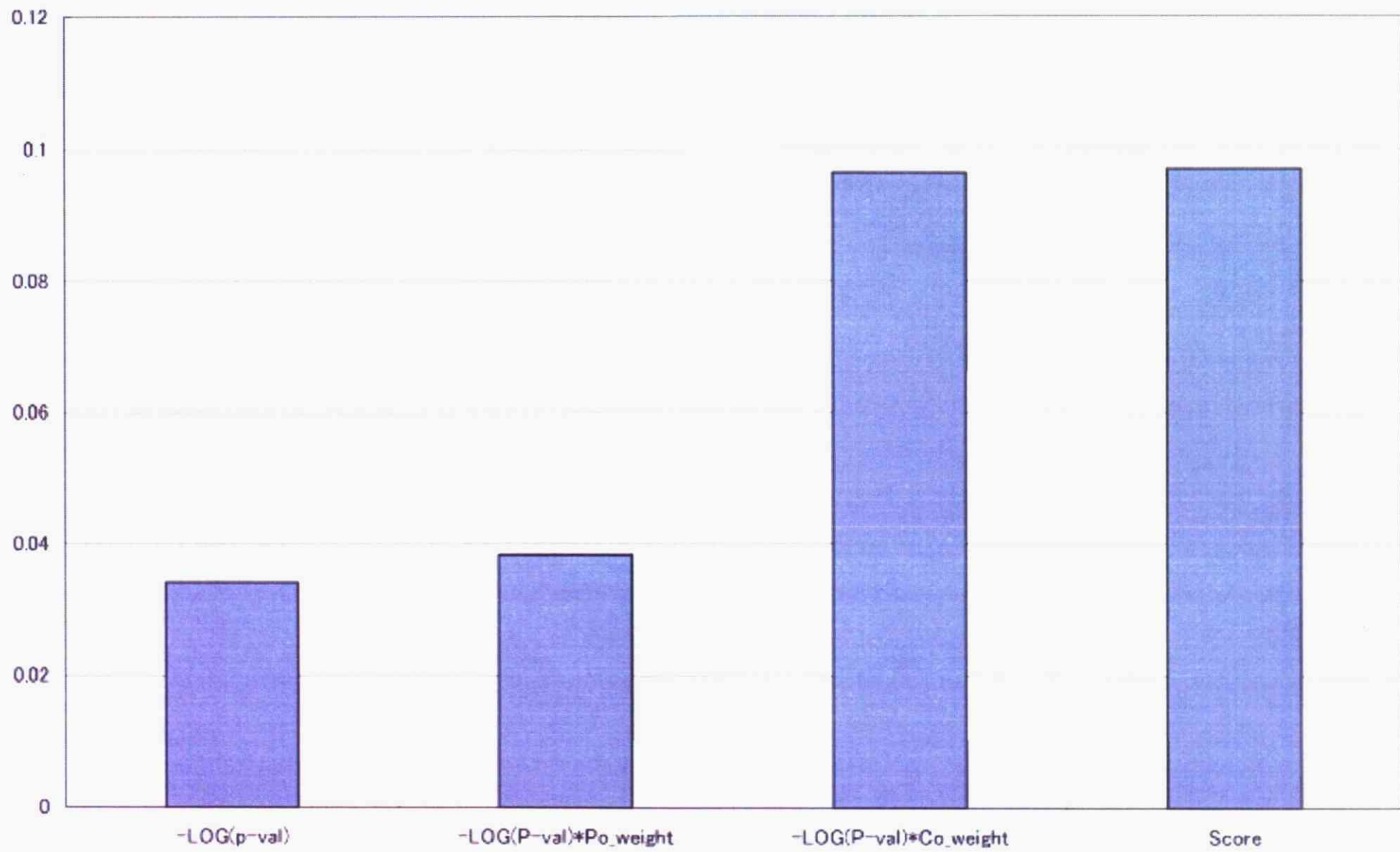


Fig. 12 重みの効果

縦軸にPeak\_INDEXの値をとり、それぞれの重みのScoreへの寄与を調べた

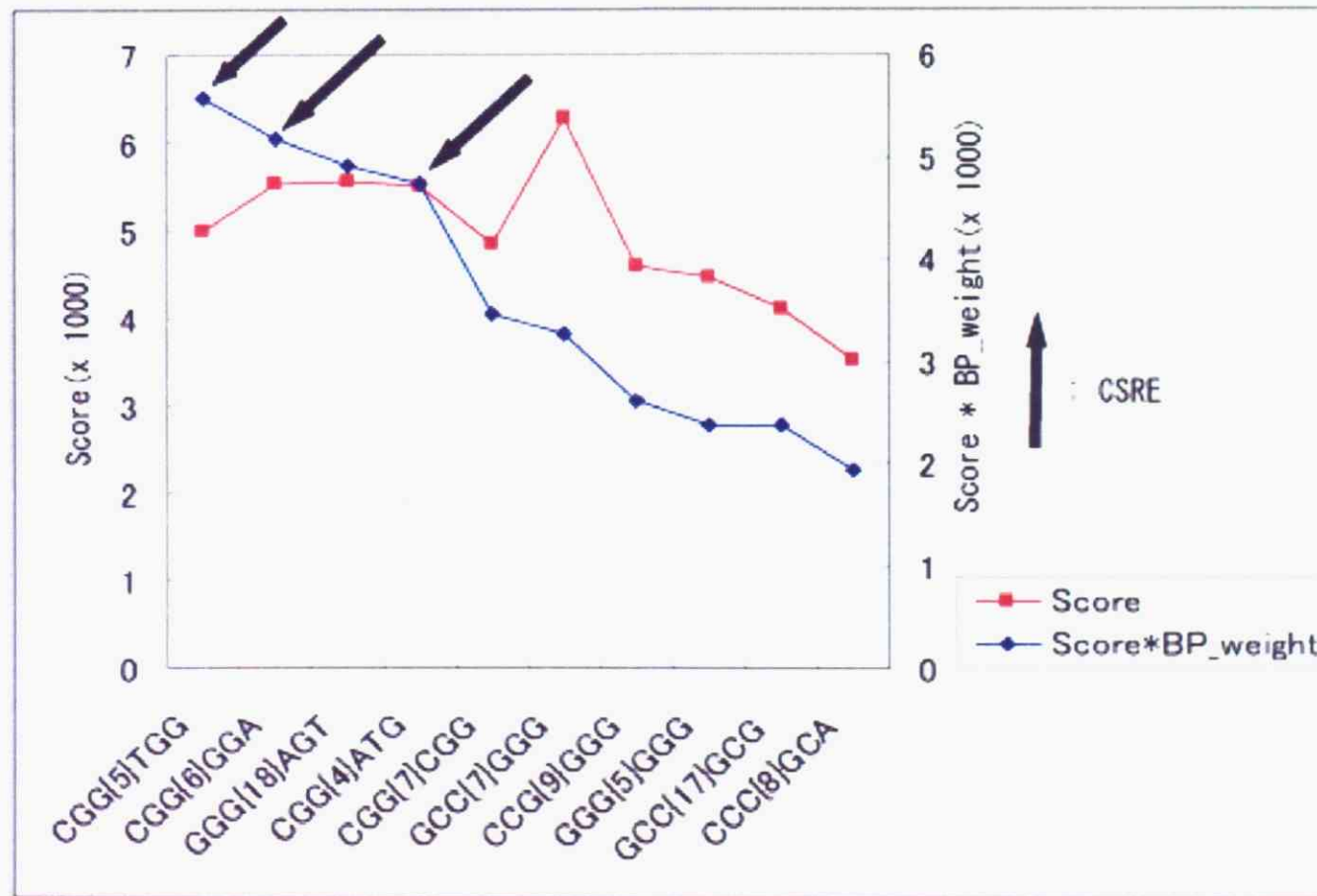


Fig. 13 BP\_weightの効果

遺伝子間の機能の類似度を評価しているBP\_weightによりCSREの順位の上昇が見られた



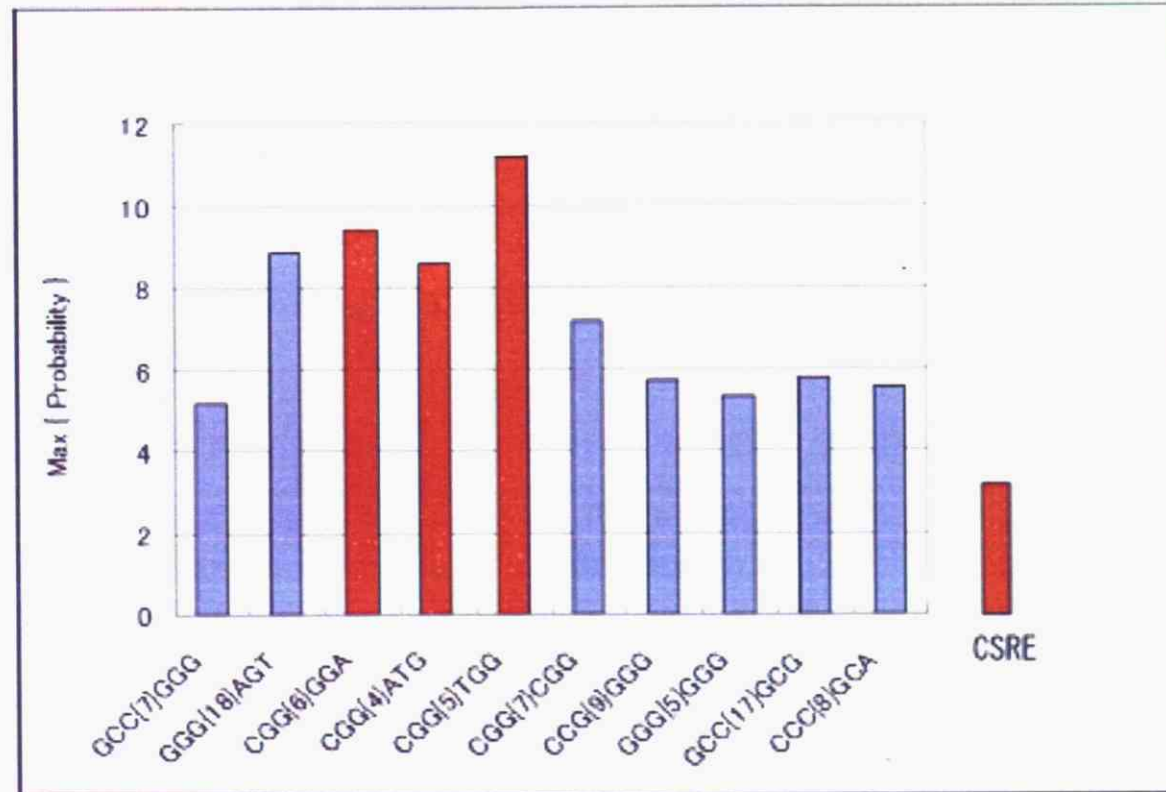


Fig. 14 各BP\_weight  
 赤で塗られているものはCSREを表している

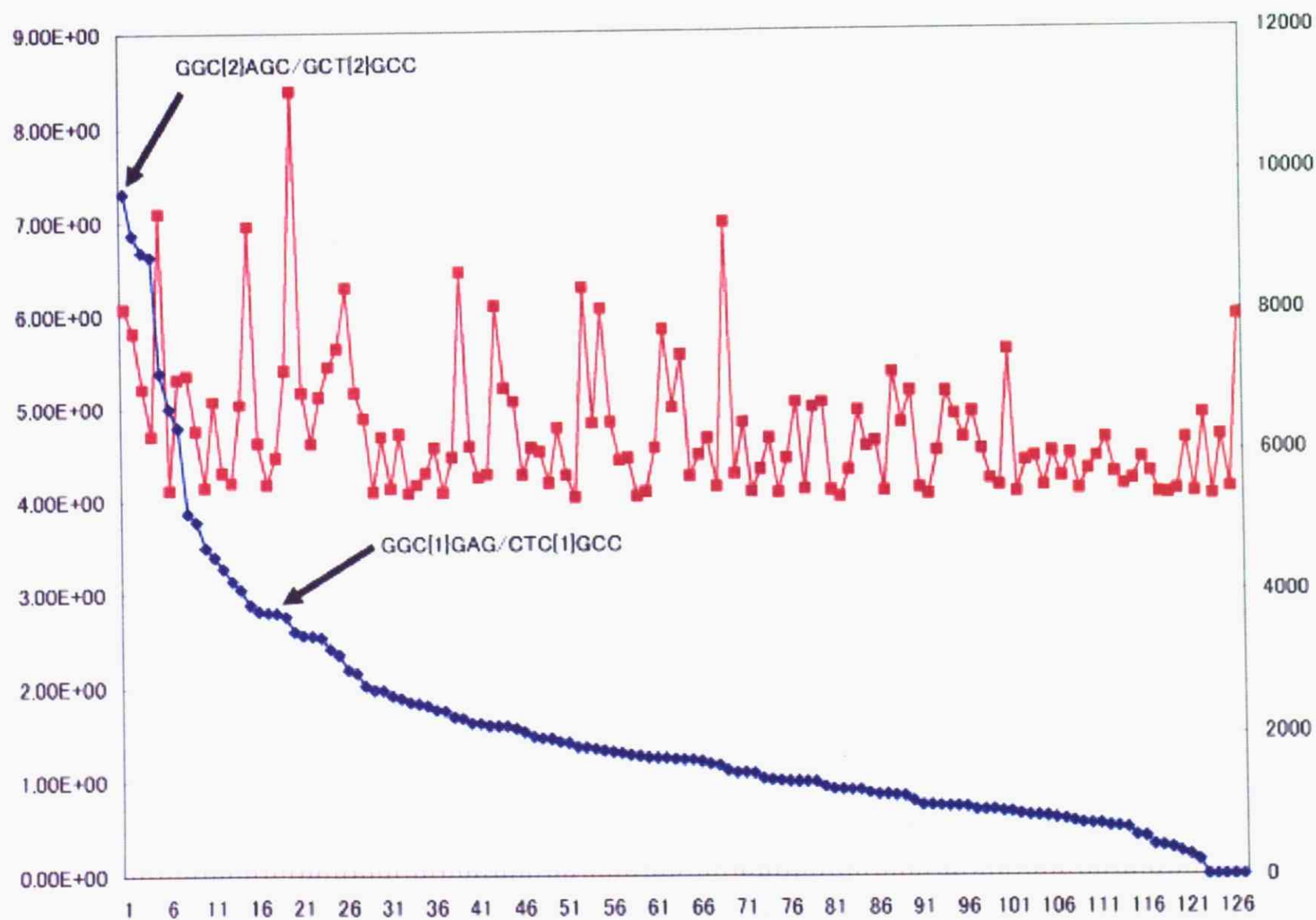


Fig. 15 コンセンサス配列既知であるPpr1にプログラムを適用した  
 左軸は $-\log(P\text{-val})$ を表し、ピンクのグラフに対応する。  
 右軸はScoreを表し、青のグラフに対応する。

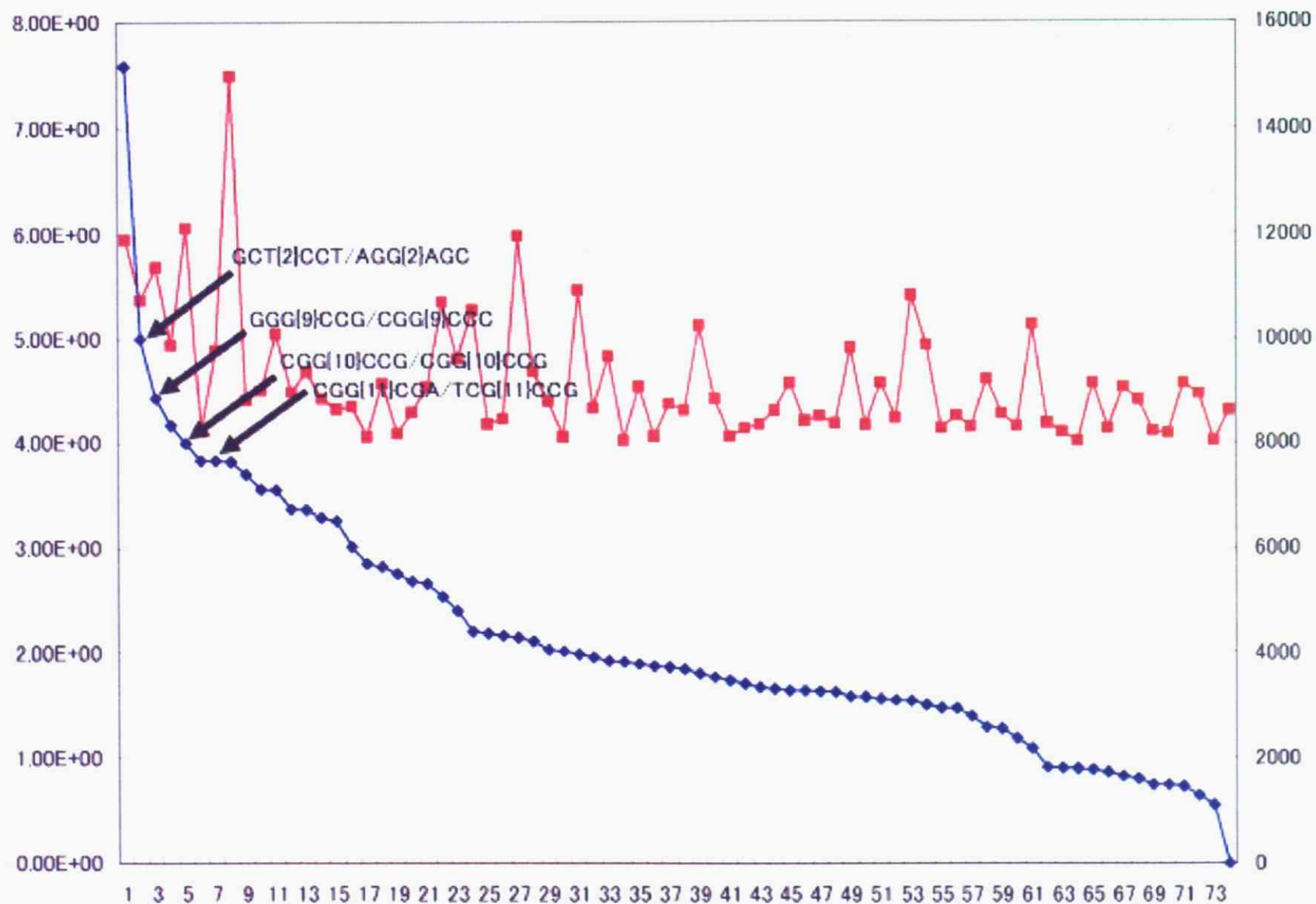


Fig. 16 コンセンサス配列既知であるPut3にプログラムを適用した  
左軸は $-\log(P\text{-val})$ を表し、ピンクのグラフに対応する。  
右軸はScoreを表し、青のグラフに対応する。  
また、黒矢印はPut3が結合すると報告されているコンセンサス配列である

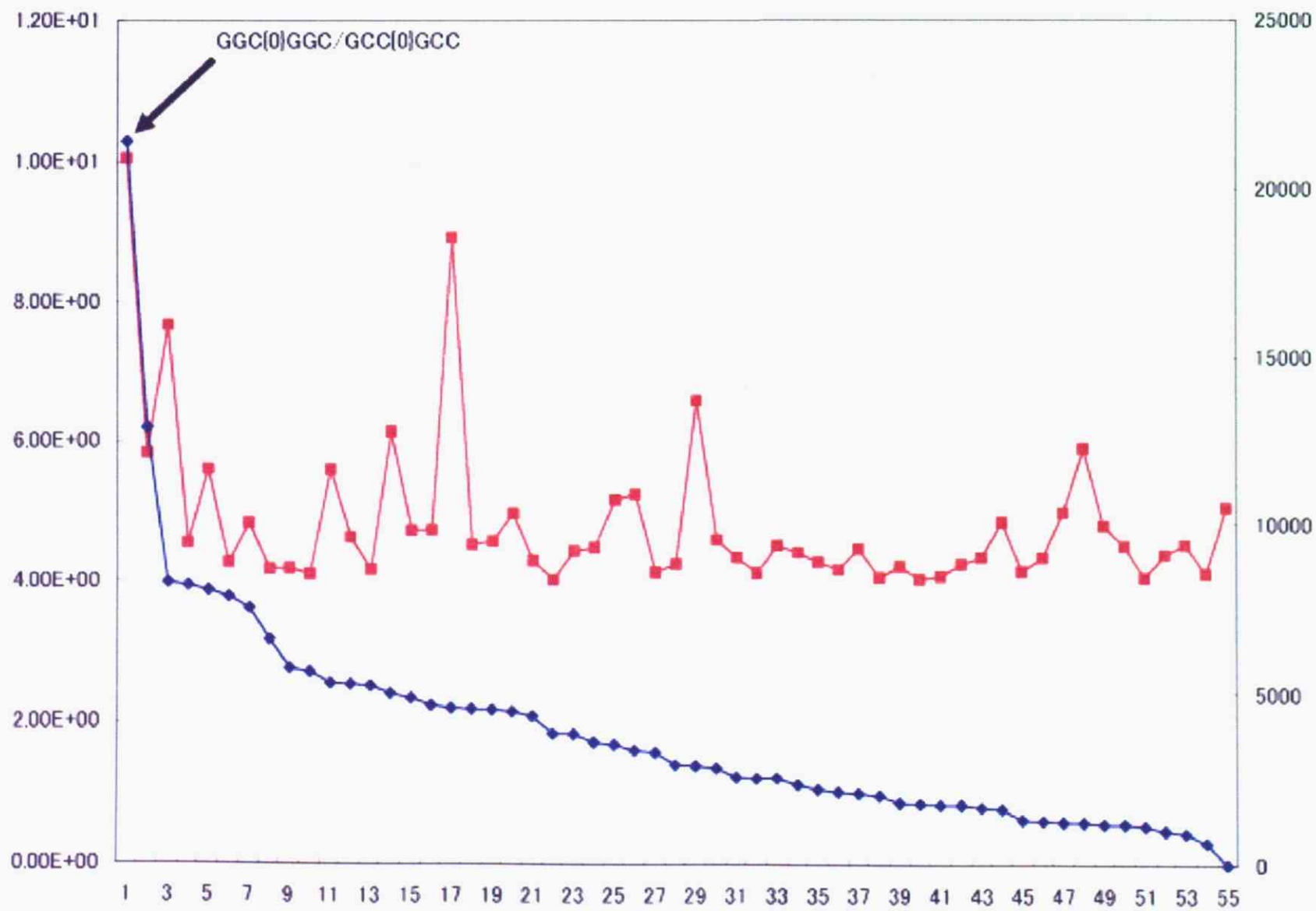


Fig. 17 コンセンサス配列既知であるUme6にプログラムを適用した  
 左軸は $-\log(P\text{-val})$ を表し、ピンクのグラフに対応する。  
 右軸はScoreを表し、青のグラフに対応する。

CCC[8]GCA/TGC[8]GGG

右軸はScoreを表し、青のグラフに対応する。

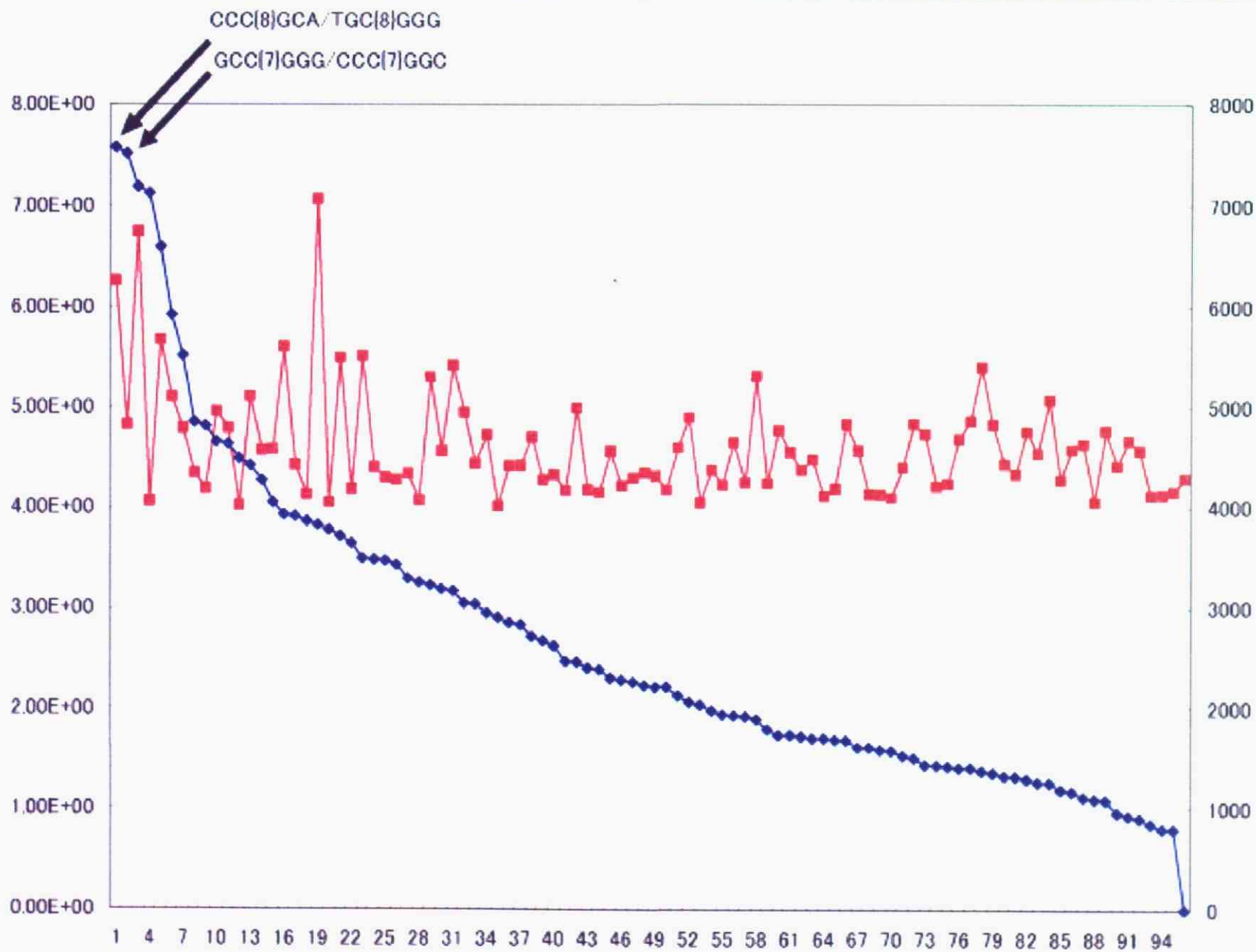


Fig. 18 コンセンサ配列が未知であるOaf1にプログラムを適用した  
左軸は $-\log(P\text{-val})$ を表し、ピンクのグラフに対応する。  
右軸はScoreを表し、青のグラフに対応する。  
また、黒矢印は高いScoreを示し、マージすることが可能な配列を示している

S p4-CF1	GTTGACCGGTGTTTTGG
S p4-CF2	ATGATTAAAGTCACCATTCAGGTTTC
pCUF	AGGGTTTTCCAGTCACGAC
S p4-PR2	TGTATCATCGAGGGAATAAGACCTGCCATATTTCTCTTGGCCATTATAGTTTTTTCTCC
S p4-Lam bdaF	CAAATATGAATGCGAGTTCCAGCGCCAATGCAAATTCTAATTTGCGGGTGGGGTGGCTCGGCAGGGATGTTCTC
pVP16	ATGTCCAGATCGAAATCGTC

\* 全て 5' -3' で表記している

表 . 1 本研究での使用プライマー



Name	ID	ratio
PCK1	YKR097W	21.31
CLI	YER065C	19.89
YAT2	YER024W	18.75
YGL157W	YGL157W	16.69
CAT2	YML042W	16.46
YAT1	YAR035W	12.00
BTN2	YGR142W	10.10
YAR040C	YAR040C	9.58
YLR154C	YLR154C	8.01
YAR037W	YAR037W	7.83
CDA1	YLR307W	7.67
YPL034W	YPL034W	7.59
DCP1	YOL149W	7.21
FBP1	YLR377C	6.84
GRE2	YOL151W	6.84
SFC1	YJR095W	6.77

MDH2	YOL126C	6.73
YDL241W	YDL241W	6.68
YLR004C	YLR004C	6.32
AQR1	YNL065W	6.29
YNR064W	YNR064C	6.06
MLS1	YNL117W	6.00
ATO3	YDR384C	5.93
SP11	YER150W	5.80
HXT12	YL170W	5.50
FRG13	YML126C	5.39
YDR222W	YDR222W	5.25
CIT2	YCR005C	5.11
YOR338W	YOR338W	4.90
BEM1	YBR200W	4.86
YLR312C	YLR312C	4.84
YGR050C	YGR050C	4.65

YOR044W	YOR044W	4.62
RK11	YOR095C	4.56
YLR311C	YLR311C	4.49
YML131W	YML131W	4.40
	YNL193W	4.36
YPL222W	YPL222W	4.28
UBC8	YEL012W	4.25
YML005W	YML005W	4.20
YPR1	YDR368W	4.11
SOL1	YNR034W	4.09

:CSRE (CCRTYSRNCCG)  
 :CSRE 1 mismatch  
 :CSRE 2 mismatches

ratio : 対照細胞と比べた発現強度

表.2 キメラ化 Sip4 により発現が誘導された遺伝子

Name	ID	Rate*
YAT2	YER024W	76.92%
PCK1	YKR097W	69.23%
FBP1	YLR377C	69.23%
CLI	YER065C	69.23%
MLS1	YNL117W	61.53%
SFC1	YJR095W	61.53%
CAT2	YML042W	61.53%
MDM2	YOL126C	53.84%
YAT1	YAR035W	46.15%
YLR311C	YLR311C	46.15%
YNL193W	YNL193W	46.15%
YLR312C	YLR312C	46.15%
YGR050C	YGR050C	38.46%
YLR004C	YLR004C	38.46%
YAR040C	YAR040C	38.46%
BTN2	YGR142W	38.46%

SPI1	YER150W	38.46%
YAR037W	YAR037W	38.46%
YML131W	YML131W	38.46%
HXT12	YL170W	38.46%
YDR222W	YDR222W	30.76%
YOR338W	YOR338W	30.76%
ATO3	YDR384C	30.76%
RNH203	YLR154C	30.76%
YPL222W	YPL222W	30.76%
DCP1	YOL149W	30.76%
YPL034W	YPL034W	30.76%
UBC8	YEL012W	30.76%
YGL157W	YGL157W	23.07%
BEM1	YBR200W	23.07%
YOR044W	YOR044W	23.07%
AQU1	YNL065W	23.07%

TRIM12	YML005W	23.07%
CDA1	YLR307W	15.38%
GRE2	YOL151W	15.38%
YDL241W	YDL241W	15.38%
RK11	YOR095C	15.38%
ERG13	YML126C	15.38%
YNR064C	YNR064C	15.38%
CIT2	YCR005C	7.69%
YPR1	YDR368W	0%
SOL1	YNR034W	0%

:CSRE (CCRTYSRNCCG)  
 :CSRE 1 mismatch  
 :CSRE 2 mismatches

\*Rate : トップクラスターに現れる確率

表.3 クラスタリングを用いた標的遺伝子濃縮