

平成23年度 森泰吉郎記念研究振興基金  
研究成果報告書

## 最小ゲノム構築に向けた最適な転写因子数の推定

### 要旨

システム生物学における挑戦のひとつに最小ゲノムを構築することが挙げられる。最小ゲノムを構成する遺伝子に関する議論はこれまでに多くなされており、バクテリアにおいては遺伝子数にして約250±100個程度と考えられている。これらの議論において、セントラルドグマ周辺のプロセスや中心代謝が考慮されやすい一方で、遺伝子発現制御に関わる転写因子は考慮されない傾向にある。これを受け本研究では最小ゲノムにおいて必須となる転写因子数の推定を行なうことを目的とする。本研究において最小ゲノムを構築する現実的なアプローチ方法は、既存のゲノムから不要な領域を削るトップダウン型であると考えられる。既存のゲノムは細胞システムを維持するために必要な構成要素で満たされているため、ゲノムの構成要素を少なくすると細胞増殖速度が落ち、最終的には細胞システムを維持することが不可能になることが予測される。したがって、細胞の生存を保障する程度の転写因子数が必要となる。本研究においては、大腸菌のゲノムスケール代謝モデル*iJR904*と遺伝子発現量を組み合わせることで、転写因子数を単純化を行なう。転写因子数が単純化されることで遺伝子発現ならびにタンパク質発現が単純化され、代謝ネットワークにおける代謝流束が単純化される。単純化した発現量を踏まえ、代謝流束解析を用いてゲノムスケールモデルにおける代謝フラックスの定量的なシミュレーションを行った。シミュレーションの結果得られたバイオマスを細胞の生存保証の閾値として用い、最小ゲノムにおける転写因子数の推定を行った。解析の結果、一定のバイオマスが保証した転写因子あたりの遺伝子は13個であった。これを遺伝子250個から構成される最小ゲノムに適応すると、転写因子20個によって最小ゲノムが制御可能となる。

キーワード：最小ゲノム、ゲノムスケールモデル、代謝流束解析、酵素反応速度論

慶應義塾大学大学院政策・メディア研究科  
池上慶太

## 1. 序論

### 1.1. 最小ゲノム

細胞が生存するために必要な最小限の遺伝子から構成されたゲノムのことを最小ゲノムとする考え方がある (Koonin, 2000). 最小ゲノムの実証的な研究は, *Bacillus subtilis* (枯草菌) を用いた実験的な最小ゲノムサイズの予測 (Itaya, 1995), *Haemophilus influenzae* (インフルエンザ菌) と *Mycoplasma genitalium* (マイコプラズマ菌) の比較ゲノム解析 (Mushegian and Koonin, 1996) が 1990年代半ばから始まった.

最小ゲノムが具体的にいくつの遺伝子数から構成されるかは, 生育条件等によって細胞が必要とする機能が異なるため, 決定することが困難である. 現状, 最小ゲノムにおける遺伝子数を決定することは困難であるものの, 最小ゲノムをどのように構築するかという観点から, 遺伝子数を類推することは可能であると考えられる. 最小ゲノム構築に向けたアプローチ方法を大分するとトップダウン型とボトムアップ型が挙げられる (O'Malley, *et al.*, 2008). トップダウン型のアプローチによる最小ゲノムの構築とは, 既存のゲノムをもとに細胞の生存に不可欠な必須遺伝子領域のみを残したゲノムを構築することを指す. この場合, 自活するバイクテリアで最小のゲノムサイズをもつマイコプラズマ菌の必須遺伝子381個が最小ゲノムになるのではないかと考えられている (Glass, *et al.*, 2006). 一方, ボトムアップ型のアプローチによる最小ゲノムの構築法は2種類存在し, 原始細胞の化学系として生命らしさを再現できるゲノム構築 (Forster and Church, 2006) と, 生体内における機能をモジュール化しゲノムを再構築する方法が存在する (Moya, *et al.*, 2009). 原始細胞系の場合, ForsterとChurchによって151個の遺伝子によって自己複製, 細胞分裂を表現できる可能性が示唆されている. また, モジュール化の場合, ノックアウト実験から *Escherichia coli* (大腸菌) における遺伝子296個 (Baba, *et al.*, 2006), 枯草菌における遺伝子271個 (Kobayashi, *et al.*, 2003) が最低限必要であることが同定されている (Juhas, *et al.*, 2011). 以上のことから, 最小ゲノムを構成する遺伝子数はおよそ250±100個ではないかと推測できる.

### 1.2. ゲノムスケールモデル

ゲノムスケールモデルとは遺伝情報の総体であるゲノム情報にもとづいて代謝ネットワークを再構成したモデルであり, 細胞内の代謝反応を定量的に表すひとつの表現方法である. 最小ゲノムを構築するにあたりゲノムスケールモデルを用いたアプローチも試みられている. 具体的には細胞内共生バクテリアのようにゲノムサイズが縮小したバクテリア (Gosalbes, *et al.*, 2010) における代謝ネットワークをシミュレーションした研究が挙げられる (Pal, *et al.*, 2006). Palらはモデル細胞内共生バクテリアのひとつである *Buchnera aphidicola* について2億年間の進化の過程で縮小化した代謝ネットワークと, ゲノムスケールモデルを代謝流束解析によって縮小化した代謝ネットワークを比較し *B. aphidicola* の代謝ネットワークの約80%を再現した. さらに, ゲノムスケールモデルと系統的クラスタリングデータを組み合わせることで, 代謝ネットワークが欠失する順序をシミュレーションした例も示されている (Yizhak, *et al.*, 2011).

### 1.3. 最小ゲノム構築に向けた課題

仮りに, バクテリアゲノムにおける遺伝子1個の平均的な塩基長を1,000塩基対とすると, 最小ゲノムが前述の遺伝子250個から構成されるとするとゲノムサイズは250kbp程度となる. この250kbpのDNAは技術的にグローニングが可能であることが示されている (Itaya, *et al.*, 2008). 別の報告ではGibsonらは7kbの化学合成したDNA断片から583kbpの機能するマイコプラズマ菌ゲノ

ムのクローニングに成功している (Gibson, *et al.*, 2008) ことから、同様に250kbp程度のゲノムを化学合成、クローニングすることは可能であると考えられる。

ただし、最小ゲノムをスクラッチから合成、クローニングできることと合成したゲノムをもつ生物が生存可能であるか否かは分けて考える必要がある。上記の研究例2件は、すでに生物としてのシステムが完成されているゲノムを対象としたゲノム全体のクローニングの成功例である。最小ゲノムが自然界において発見されない場合、少なからず人工的にゲノムを合成する必要がある。その際、課題となるのは人工的に合成したゲノムが生命としての機能をもつか否かであると考えられる。人工的に合成したゲノムの例ではないが、ゲノムを機能的モジュールの集合であると捉えた場合に、機能的モジュールである遺伝子回路を既存の生命システムに実装することの難しさが報告されている (Purnick and Weiss, 2009)。Purnickらの報告によると、2000年以降、遺伝子回路を実装した研究例は増えている一方で遺伝子回路の規模自体は頭打ちされていると指摘する。また、彼らは遺伝子回路の規模をプロモーター数と定義しており、遺伝子回路における遺伝子発現調節の制御の難しさ、ひいては遺伝子回路が生体内において機能することの難しさを指摘する。

#### 1.4. 本研究の目的

先に述べた通り、現状考察が行われている最小ゲノムにおいては転写因子が考慮されていない。細胞内で機能するゲノムを合成するためには遺伝子の発現制御が不可欠である。これを受け本研究では最小ゲノムを構築するために必要な転写因子数の推定を行なう。最小ゲノムにおいては構成要素を最大限少なくし、転写因子数もまた最大限少なくする。また、最小ゲノムを構築する現実的なアプローチ方法は、既存のゲノムから不要な領域を削るトップダウン型を採用した。既存のゲノムは細胞システムを維持するために必要な構成要素で満たされているため、ゲノムの構成要素を少なくすると細胞増殖速度が落ち、最終的には細胞システムを維持することが不可能になる (Mizoguchi, *et al.*, 2008)。したがって、細胞の生存を保障する必要がある。本研究においては、大腸菌のゲノムスケール代謝モデルを用いることで、転写因子数を単純化をモデルする。転写因子数が単純化されることでタンパク質発現が単純化され、代謝流束もまた単純化することで、定量的なシミュレーションを行い、バイオマスを計算することで細胞の生存を保証する (図1.1)。

第2章では、遺伝子発現にもとづく代謝フラックス導出のためマイクロアレイデータから反応速度を求める方法について議論を行なう。まず、(i) 細胞内における mRNA量とタンパク質存在量の相関関係から回帰式を導出することで、マイクロアレイデータをタンパク質存在量へ変換することが可能となる。次に、(ii) 大腸菌のゲノムスケール代謝モデルを用いて代謝流束解析からバイオマスを最大化するフラックス分布を求める。そして、(iii) 酵素反応速度論より反応速度式を計算する。(iii)において反応速度式に特異的な定数を(i)と(ii)を用いて解く。また、(iii)における反応速度は(i)における回帰式を用いて、マイクロアレイをタンパク質存在量に変換して計算する。

第3章では第2章で得られたデータをもとに最小ゲノムにおける転写因子の数について議論を行なう。まず(iv)ゲノムスケールモデルの遺伝子をグループ化し、グループごとに遺伝子発現量を離散化する。このとき、グループ1個が転写因子1個であると仮定する。次に、(v) 離散化した遺伝子発現量をもとに、(i)の回帰式を用いてタンパク質存在量へ変換し、さらに(iii)を用いて代謝ネットワークの全反応速度を求め、バイオマスを計算する。そして、グループ化はグループに含まれる遺伝子数を昇順に増やしていき、そのつどバイオマスを計算することによつ

て、遺伝子数1個からゲノムスケールモデルの遺伝子すべてが含まれるまで連続的にバイオマスの推移を解析することが可能となる。(vi) 野生株に対して一定のバイオマスを保障するグループ化は、細胞の生存を保障する発現量の分布であるとみなし、最小ゲノム構築に必要な転写因子数であると推測する。

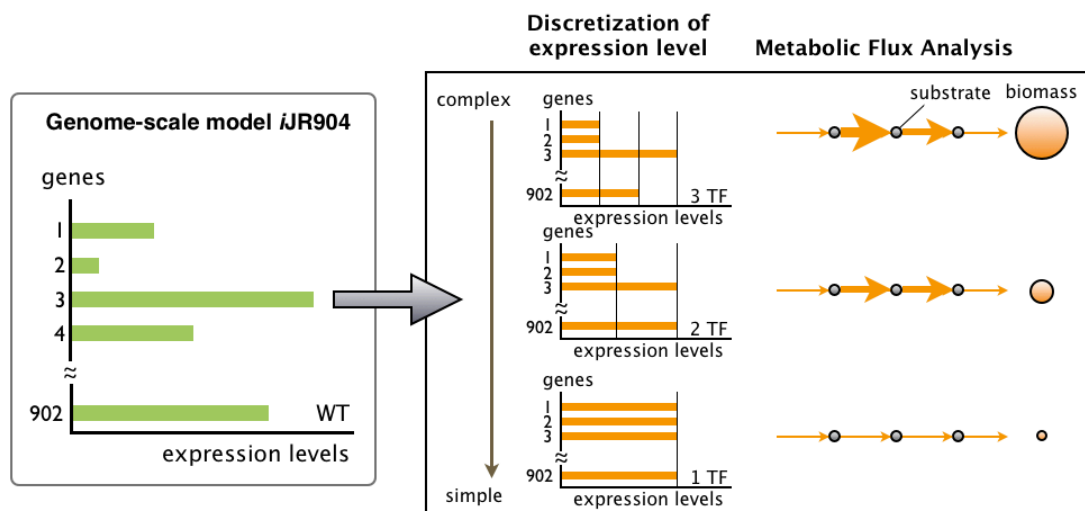


図1.1: 本研究の概略

最小ゲノム構築に向けた最適な転写因子数の推定を行なうため、遺伝子発現量を離散化しバイオマスを計算する。本研究で使用したゲノムスケールモデルは904個の遺伝子から構成されており、このうち遺伝子2個に関してはアノテーションが不足しているため本解析には用いていない。マイクロアレイで測定された遺伝子発現量データを使用する。発現量の離散化はグループに含まれる遺伝子数1個から昇順で902個まで連続的に行い、それぞれの離散化パターンでバイオマスを計算する。グループ1個が転写因子1個に相当し、同時に同程度の発現強度をもつと仮定し、バイオマスの推移にもとづいて転写因子数を単純化する限界を探索する。なお、バイオマスの閾値は野生株におけるバイオマス95%まで保証するものとする。

## 2. 遺伝子発現にもとづいた代謝流束の導出

### 2.1. 背景

代謝流束解析を用いた先行研究において、遺伝子の発現量をモデルに組み込んでいる例は存在するものの、これらの場合において発現量を代謝流束解析における解空間の制約条件として用いているため、必ずしも発現量を考慮した代謝フラックスを計算しているとは言いがたい (Colijn, *et al.*, 2009)。そこで、今回新たな手法として、本研究では酵素反応速度論において反応速度を代謝フラックスと発現量から導出する式変形を行った。酵素反応速度論における酵素濃度はマイクロアレイを用いて測定した遺伝子発現量データをもとに求めた。本章ではマイクロアレイ発現量データから代謝流束解析にもとづいた反応速度を計算方法を導出することを目的とする。

本研究において、現在質量分析器を用いたタンパク質の定量手法は発展しつつあるものの、マイクロアレイやシーケンサーを用いたmRNAの定量手法と比較すると未だにタンパク質の絶対定量には課題が多い。そこで、mRNA量からタンパク量の推定方法が確立されれば、タンパク質の存在量が比較的容易に測定可能となる。細胞内におけるmRNA量とタンパク質存在量は比例関係にある前提で解析を行なう。

代謝流束解析は代謝ネットワークにおける流束の分布を定量的に解析する手法のひとつである (Edwards and Palsson, 1998). 代謝フラックスの分布を求めることは、生体内の代謝ネットワークにおける代謝反応がどのような代謝経路でどの程度行われているかを知ることにつながる。代謝流束解析によるシミュレーションによって最適化された代謝フラックスの予測分布は、実験的に測定された実測値に相関があることが先行研究によって示されている (Blank and Kuepfer, 2010).

## 2.2. 対象と手法

### 2.2.1. 細胞内における発現量とタンパクの相関関係

この結果を受け、Gene Expression Omnibus (GEO) (Barrett, *et al.*, 2011) に登録されている大腸菌のマイクロアレイデータを網羅的に取得し、mRNA 発現量とタンパク質存在量の相関関係を解析した。GEOより、大腸菌K-12株、野生型、pH7の条件に該当するすべてのマイクロアレイデータ2669個を取得し、解析の対象とした。この中からIshihamaらによって大腸菌K-12株(野生型)のタンパク質がLC-MS/MSによって定量的に測定されたデータ (Ishihama, *et al.*, 2008) (N = 990) と相関が最も高いマイクロアレイデータの探索を行った。最も相関の高かったマイクロアレイデータとIshihamaらのタンパク質存在量データでは遺伝子986個が共通する。このうち遺伝子4個はタンパク質が同定されているもののマイクロアレイデータでは測定できないため、相関解析から除外した。

### 2.2.2. 酵素反応速度論を用いた反応速度式の構築

本研究に用いる代謝流束解析のモデルとして、大腸菌のゲノムスケールモデル代謝モデルのひとつである*iJR904* (Reed, *et al.*, 2003) を選択した。*iJR904* には遺伝子904個、代謝反応931個、代謝物625個が定義されている。代謝反応数に関してはアイソザイムの存在によって、ひとつの代謝反応名に対して複数の反応が定義されている場合がある。したがって、*iJR904*では931個の反応が定義されているものの、重複する代謝反応を考慮し、実際の代謝反応数は1061個が存在する。また、交換代謝流として交換代謝反応143個が定義されている。代謝流束解析を行なうにあたり、グルコースの取り込み速度は10[mmol/gDW/h]に、酸素の取り込み速度は20[mmol/gDW/h]に設定した。

代謝流束解析のシミュレーションに用いる酵素反応速度*V*はミカエリス・メンテン式より、酵素反応速度*V*を得る。このとき、 $V_{\max}$ は最大反応速度、 $K_m$ はミカエリス定数、 $[S]$ は基質濃度、 $[E]$ は酵素濃度を表す。また、最大反応速度 $V_{\max}$ は

$$V_{\max} = k_{\text{cat}}[E]$$

と表すことが可能である。 $k_{\text{cat}}$ は回転数を表す。これをミカエリス・メンテン式に代入する。

$$V = k_{\text{cat}}[S][E] / (K_m + [S])$$

ここで、 $k_{\text{cat}}$ 、 $K_m$ 、 $[S]$ の定数性を仮定して

$$\alpha = k_{\text{cat}} / (K_m + [S])$$

と置き、反応速度*V*を得る。

$$V = \alpha[S][E]$$

### 2.2.3 仮説の検証

マイクロアレイの発現量データをもとに代謝フラックスを計算するための反応速度式を構築する過程で仮定した $\alpha[S]$ の定数性について以下の3種類の検証を行った。

検証 I) ミカエリス・メンテン式より導出した反応速度式 $V$ について、定常状態下において定数的な振る舞いをするを仮定した $\alpha[S]$ について以下の検証を行う。試験群の $\alpha[S]$ はグルコースを炭素源としたときのシミュレーションを行った。対照群の $\alpha[S]$ は、6種類の炭素源をそれぞれ独立の培地で培養したときの発現量変化をマイクロアレイで測定したデータを使用した (Liu, *et al.*, 2005)。これによって、外部変動に対して $\alpha[S]$ が摂動を受けないことを検証する。 $\alpha[S]$ は、反応速度 $V$ を $\alpha[S]$ について解くことによって得ることが可能である。なお $V$ はゲノムスケールモデルにおける反応速度を表す。 $[E]$ は酵素濃度を表し、Liuらのマイクロアレイデータを用いた。Liuらは、大腸菌において炭素源となる6種類の有機化合物(グルコース、グリセロール、酢酸、コハク酸、アラニン、プロリン)を唯一の炭素源として最小培地で培養したときの発現量の変化をマイクロアレイで測定した (N=15)。

酵素濃度 $[E]$ は、2.2.1より求めた回帰式を用いて、Liuらの発現量データを酵素濃度として変換した。また、反応速度 $V$ は代謝流束解析よりシミュレーションし、炭素源6種類に関して取り込み速度は先行研究より参照した (Fong and Palsson, 2004; Meadows, *et al.*, 2010)。すべての実験は好気条件下で行われたため酸素の取り込み速度は統一的に20[mmol/gDW/h]とした。 $\alpha[S]$ の比較には、試験群と対照群 (N = 15) の平均差の検定をt検定 (対応あり) を用いて行った。

検証 II) 反応速度パラメーター $K_m$ 、 $[S]$ はChassagnole らのモデルを参照した (Chassagnole, *et al.*, 2002)。Chassagnoleらは大腸菌においてグルコースからアセチルCoAまでの解糖系とペントースリン酸経路における詳細なモデルを表現し、交換代謝を含む関連する代謝反応34個をモデル中で定義した。ゲノムスケールモデルにおけるすべての代謝反応のうち、Chassagnoleらのモデルで定義され、かつ反応速度パラメーターが揃っている代謝反応を対象として検証を行った。なお、 $k_{cat}$ に関しては反応速度パラメーターデータを参照した。 $\alpha[S]$ の比較は、検証Iと同様に、試験群と対照群(N = 15)の平均差の検定をt検定(対応あり)を用いて行った。試験群の $\alpha[S]$ は検証Iで示した通りの条件でシミュレーションした。一方、対照群の $\alpha[S]$ は各反応速度パラメーターを先行研究より参照した。より詳細な反応速度パラメーターから $\alpha[S]$ を計算し、試験群の $\alpha[S]$ と比較することで $\alpha[S]$ の定数性を検証した。

検証 III) Chassagnole らによって構築された代謝反応の速度式に反応速度パラメーターを代入し反応速度 $V$ を計算した。代謝反応 34 個のうち交換代謝を除く23個を対象とし、検証Iと同様の条件でシミュレーションし、 $\alpha[S]$ が含まれる反応速度 $V$ と、前述の定常条件下における詳細な代謝モデルであるの反応速度 $V$ の比較検証を行った。

## 2.3. 結果

### 2.3.1. 細胞内における発現量とタンパクの相関関係の検証

各マイクロアレイデータとタンパク質存在量における2669対のデータをピアソン相関係数で確認した。その結果、GSM101237が最も高い相関 ( $R^2 = 0.37$ ) を示した (図 2.1)。

マイクロアレイデータ GSM101237 と Ishihama らのタンパク質存在量データから回帰式を割り出した。

$$Y \approx 2.0333 \times X^{0.8037421}$$

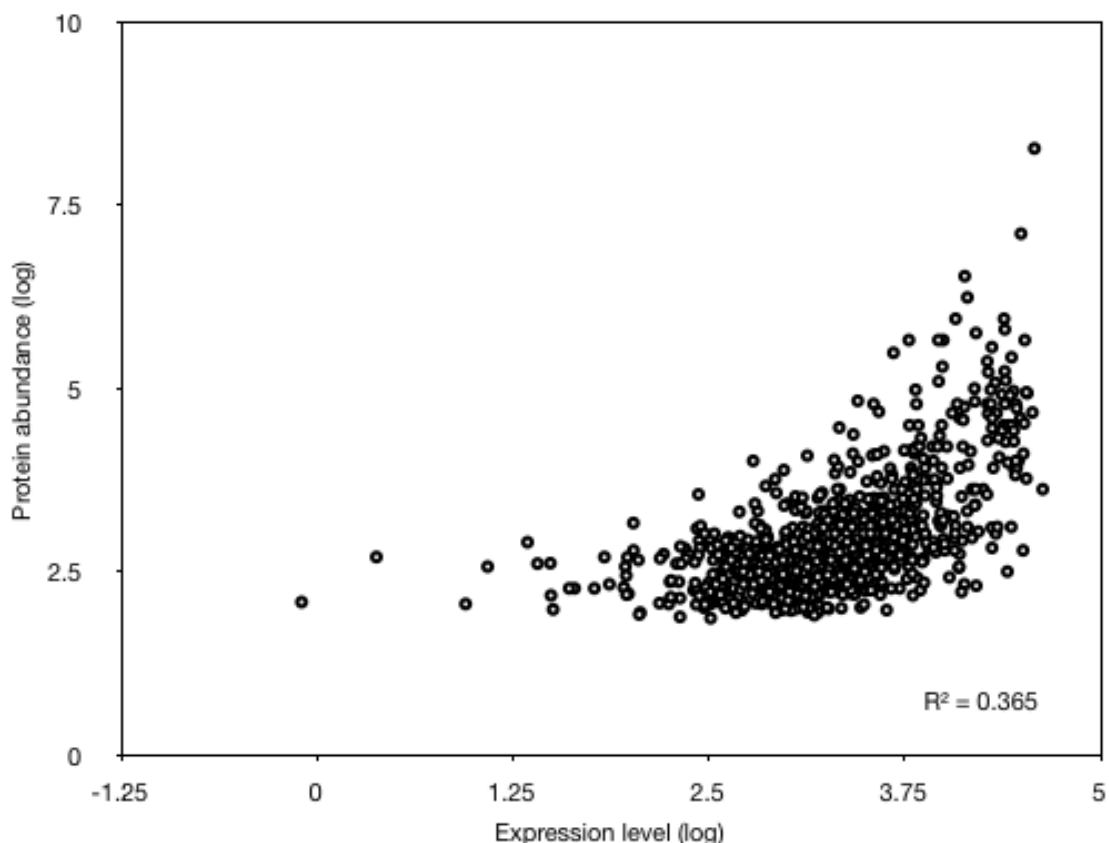


図2.1: マイクロアレイデータとタンパク質存在量データの比例関係

最もタンパク質存在量データの当てはまりのよかったマイクロアレイサンプル GSM101237(GPL199) における散布図。縦軸はタンパク質存在量 ( $\log_{10}$ スケール), 横軸はマイクロアレイ発現量 ( $\log_{10}$  スケール) を表す(ピアソン相関係数  $R^2 = 0.365$ ). マイクロアレイとタンパク質存在量データに共通した遺伝子の割合はタンパク質存在量データに対して約99.6%である( $N = 986$ ).

### 2.3.2. 酵素反応速度論を用いた反応速度式

バイオマス最大化を目的関数としたシミュレーションの結果, バイオマス  $0.92[\text{mmol/gDW/h}]$  が得られた. この際, 代謝反応931個のうち588個 (63%) がゼロフラックスであった.

### 2.3.3. 仮定の検証

検証Iの結果を示す. 各炭素源に関して代謝流束解析を用いてシミュレーションを行った. Liuらによる先行研究とシミュレーション結果を比較したところ, ピアソンの相関係数  $R^2 = 0.60$ , スピアマンの順位相関係数  $\rho = 0.79$  となった. 先行研究に対し, グリセロールのバイオマスがグルコースよりも高い原因は, グリセロールの取り込み速度がグルコースよりも速いことが起因していると考えられる. 試験群と対照群の平均差を比較したところ, 発現量の違いから代謝フラックスはそれぞれ異なった結果となったものの, 統計的な有意な差を検出できなかった. すなわち, 両群は同一のものであると統計的に解釈できる. 検証IIについて, 代謝反応34個のうち22個について反応速度パラメータ  $k_{cat}$ ,  $K_m$ ,  $[S]$  を取得することができた. 試験群と対照群の平均差を比較したところ, 統計的な有意な差を検出できなかった ( $p\text{-value} = 0.61$ ). すなわち, 両群における  $\alpha[S]$  は同一のものであると統計的に解釈できる. 検証IIIの結果を示す. 代謝流束解析ではゼロフラックスとなる代謝反応EDDとG3PD5がChassagnoleらのモデルでは代謝フラックスが確認された. 両反応速度の順位相関は  $\rho = 0.42$  ( $p\text{-value} = 0.04$ ) となった.

## 2.4. 議論

Luらと Ishihama らのタンパク質存在量に関して、Luらタンパク質存在量データとマイクロアレイ発現量データ 423 対について相関を解析した。一方、本解析では 986 対のデータを対象とした。対象となるタンパク質存在量データは本研究の方が約 2 倍のサンプル数を持つ。また、タンパク質存在量データに対するマイクロアレイデータを比較すると Luらの先行研究では 3 つのマイクロアレイデータの平均値をひとつのマイクロアレイデータとして比較しているのに対して、本研究では GEO に登録されている独立した大腸菌の全マイクロアレイサンプルデータに対する比較を行った。以上のことから、タンパク質存在量データとマイクロアレイ発現量データのフィッティングは先行研究の方が高いものの Ishihama らが測定したタンパク質存在量データとマイクロアレイサンプルデータ GSM101237 を比較した結果を本研究における両データの相関係数とし、同時に算出した回帰式を用いてマイクロアレイデータをタンパク質存在量へ変換する。 $\alpha[S]$ の外部変化として酵素濃度[E]に炭素源を制限するという条件を加えることで遺伝子の発現量に多様性を持たせ、反応速度  $V$  もまた炭素源に合わせたシミュレーション結果を用いた。解析の結果、外部変化に対して $\alpha[S]$ は一定の値を取った。このことは近年研究が進められ明らかになってきている代謝ネットワークにおける恒常性とも矛盾しない (Ishii, et al., 2007)。すなわち、環境変化にตอบสนองして遺伝子の発現やタンパク質の合成は調整される一方で代謝物質の濃度に変化は少ない。このことは、 $\alpha[S]$ の定数性に関して基質濃度  $[S]$  やミカエリス定数  $K_m$ 、最大反応速度  $V_{max}$  においても定数に近い挙動を示すことを示唆できるのではないかと考えられる。検証Iのまとめとして、 $\alpha[S]$ は外部の変化に対して摂動を受けないことを確認した。より詳細な反応速度パラメータを用いた $\alpha[S]$ の検証を次の「検証 II」において行なう。より詳細な反応速度パラメータから $\alpha[S]$ を計算した対照群と試験群の $\alpha[S]$ と比較したところ統計的な有意差を見出すことがなかった。この検証から、 $\alpha[S]$ は定数的な挙動を示すと解釈できる。本研究で用いる $\alpha[S]$ は代謝ネットワークが定常状態下において定数的な振る舞いをとると仮定した。これを検証するために 3 つの検証を試みた。検証Iにおいては多数のマイクロアレイデータに対して酵素濃度  $[E]$  を算出し、マイクロアレイデータに応じた反応速度  $V$  を計算し、検証用の $\alpha[S]$ を用意した。両 $\alpha[S]$ を比較したところ統計的に同一の値であることを示した。検証 II においては詳細な反応速度パラメータから検証用の $\alpha[S]$ を用意した。そして、両 $\alpha[S]$ を比較したところ統計的に同一の値であることを示した。検証IIIにおいては $\alpha[S]$ を用いて計算した代謝フラックスと詳細な反応速度パラメータから計算した代謝フラックスを比較した。そして、両代謝フラックスに相関があることを確認した。以上、3つの検証から本研究で用いる  $\alpha[S]$  は代謝ネットワークが定常状態下において定数的であると解釈する。

## 3. 発現量の離散化を用いた転写因子数の推定

### 3.1. 背景

バクテリア(真正細菌)における転写因子数の研究は、モデル生物である大腸菌を中心として研究が進められている (Browning and Busby, 2004)。大腸菌における転写因子数は317個が予測されており (Perez-Rueda and Collado-Vides, 2000)、実験的には175個が同定されている (Gama-Castro, et al., 2011)。本章では、大腸菌のゲノムスケール代謝モデルを用いて、転写因子を単純化することによって転写因子数を推定した。第2章ではマイクロアレイで測定した遺伝子発現量データにもとづいて反応速度  $V$  を導いた。本章では最小ゲノムにおける転写因子を単純化を表現するためにゲノムスケールモデルの遺伝子をグループ化を行なう。ひとつの転写因子によって制御



される遺伝子群をひとつのグループと定義する。グループに含まれる遺伝子数を1個から902個までのすべてのパターンで代謝流束解析を用いてシミュレーションし、最小ゲノムにおける転写因子数の最少数を推定する。

## 3.2. 対象と手法

### 3.2.1 バイオマスの計算

バイオマスの計算は、バイオマスの前駆体となる各基質について合計フラックスを計算した。任意の基質について合計フラックスを、基質の化学量論係数と基質が含まれる代謝反応の代謝フラックスの積として表し、物質収支を考慮して基質が右辺の場合は1、左辺の場合は-1を掛けあわせた。これを基質が含まれる代謝反応 $n$ 個について計算し、総和を基質の合計フラックスとした。バイオマスの前駆体53個すべてについて合計フラックスを算出し、左辺基質の合計フラックスの中央値と右辺基質の合計フラックスの中央値の平均値をバイオマスとした。なお、フラックスは第2章「マイクロアレイデータから反応速度を計算」と同様のシミュレーション条件で計算した。

### 3.2.2 発現量の離散化

ゲノムスケールモデルにおいて定義される遺伝子904個(アノテーションの関係により解析に用いる遺伝子数は902個)について、遺伝子発現量を離散化しグループ化を行ないバイオマスを計算する。野生株を除くと、すべてのグループには複数の遺伝子が含まれ、各遺伝子は固有の遺伝子発現量をもつ。グループ内の発現量の離散化は、基本統計量である最大値、最小値、最頻値、中央値、平均値を用いた。

## 3.3. 結果

### 3.3.1 バイオマスの計算

バイオマスについて、代謝流束解析でバイオマスを最大化したときの値と代数的にバイオマスの反応式から解いた値を比較した。左辺と右辺の合計フラックスの総和の中央値はそれぞれ0.911[mmol/gDW/h]、0.923[mmol/gDW/h]であった。結果として代謝流束解析と同様のバイオマスが代数的に算出されたことを確認した (0.92[mmol/gDW/h])。.

### 3.3.2 転写因子数の推定

ゲノムスケールモデルの遺伝子 902 個について連続的にグループ化し、グループに含まれる遺伝子発現量を離散化した(図3.1)。そして、離散化した遺伝子発現量をもとにバイオマスを計算した。データの線形フィッティングは $R^2=0.87$ である。野生株のバイオマスは 0.92[mmol/gDW/h] であるため閾値は 0.86[mmol/gDW/h] とした。バイオマス 95%を保証するとき、グループ数は 69 個であり、グループに含まれる遺伝子数が 13 個である。よって、95%保証のとき転写因子 1 個あたり遺伝子 13 個を制御することが示された。発現量  $\Delta$  を固定し、離散化した遺伝子発現量をもとにバイオマスを計算した (図 3.6)。データの線形フィッティングは $R^2=0.45$  である。野生株のバイオマスは0.92[mmol/gDW/h] であるため閾値は 0.86[mmol/gDW/h] とした。バイオマス 95%を保証するとき、グループ数は 178 個であり、グループに含まれる平均遺伝子数が 6 個である。よって、タイプ 2 によるグループ化は 95%保証のとき転写因子 1 個あたり遺伝子 6 個を制御することが示された。

発現量  $\Delta$  を可変とし、離散化した遺伝子発現量をもとにバイオマスを計算した。データの線形フィッティングは  $R^2=0.30$  である。野生株のバイオマスは  $0.92[\text{mmol/gDW/h}]$  であるため閾値は  $0.86[\text{mmol/gDW/h}]$  とした。バイオマス 95% を保証するとき、グループ数は 320 個であり、グループに含まれる遺伝子数が 3 個である。よって、タイプ 2 によるグループ化は 95% 保証のとき転写因子 1 個あたり遺伝子 3 個を制御することが示された。

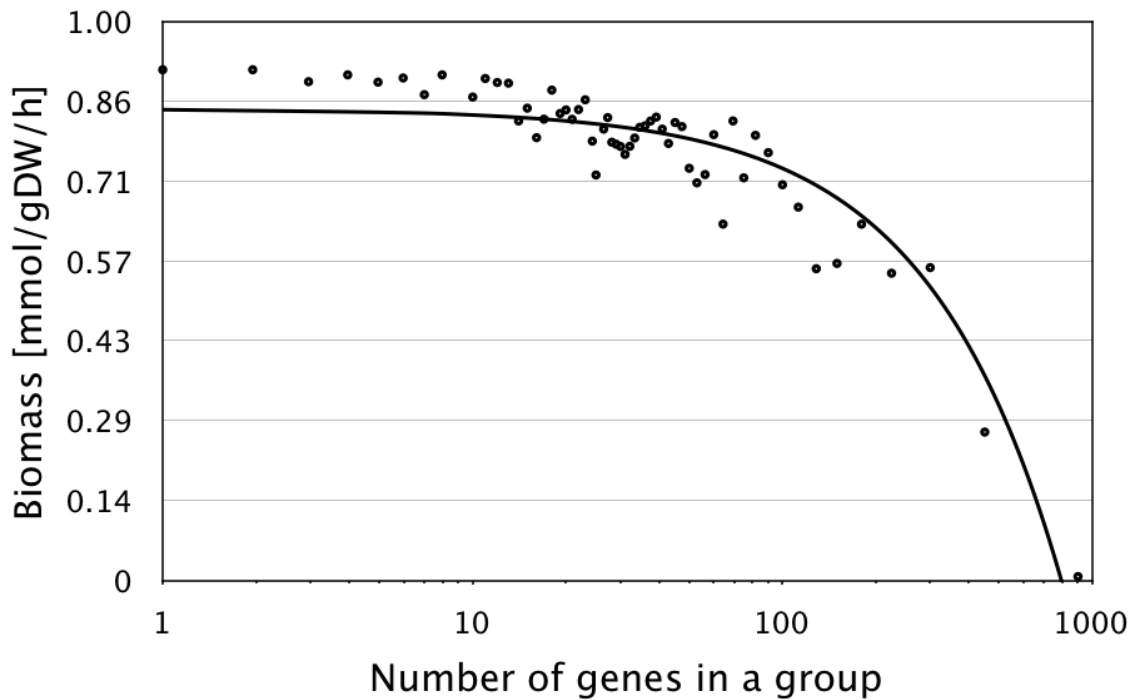


図3.1: タイプ 1 によるバイオマスの結果

縦軸がバイオマスを表し、横軸がグループに含まれる遺伝子数を  $\log_{10}$  スケールで表す。グループ数を少なくするほどバイオマスが減少する傾向にある ( $N=902$ )。

### 3.4. 議論

3 種類のグループ化の結果、グループに含まれる遺伝子数を連続的に増加させた場合が最も転写因子数を単純化することができた。大腸菌において ORF4200 個を Perez-Rueda らが予測した転写因子数 317 個 (Perez-Rueda and Collado-Vides, 2000) で割ると転写因子 1 個あたり 13 個の遺伝子を制御していることになり、タイプ 1 によるグループ化においても野生株のバイオマスを 95% 保証したときと同程度の制御効率であることが示される。最小ゲノムが遺伝子 250 個で構成されるとすると、約 20 個の転写因子を考慮する必要があることが本研究の結果から示唆できる。

## 4. 結論

最小ゲノムは「最小」という名前の通り最も小さなゲノムのことであるが、ゲノムサイズ、遺伝子数、生体内機能数といった数の多少を競争するよりもゲノムを構成する要素、その要素間の相互作用、要素が系に及ぼす影響について人間の理解が及びやすいことに意義があると考え

える。この意味において、ゲノムが最小であることはゲノムの構成要素も少なくなり、相対的にゲノムが大きいものよりも理解が進むのではないかと思う。また、最小ゲノムは遺伝子回路、遺伝子回路を構成するパーツといったモジュール (Smolke, 2009) を組み込むことによって、多様な機能を加えたときの系の振る舞いを観察することができるプラットフォームを提供できるのではないだろうか (Ellis, *et al.*, 2011)。ひいては、有用物質を特異的かつ効率的に生産することを可能にする (Reaves and Rabinowitz, 2010)。以上より、最小ゲノムを構築する意義は、細胞システムをより理解するため、ゲノムの構成要素が明らかとなっているものを合成することであり、機能モジュールのプラットフォームとしての役割にあると考える。

バクテリア型の細胞を人工的に合成するためには遺伝情報の総体であるゲノムの合成と、ゲノムを他の系から区切る細胞膜といった容器の合成が不可欠である (Szostak, *et al.*, 2001)。そして本研究は合成したゲノムが細胞内で機能的に働くために欠かすことのできない転写因子数を推定することを目的とした。この目的を達成するために、大腸菌のゲノムスケール代謝モデルを用いて、第2章ではマイクロアレイから反応速度を求め、第3章において3種類のグループ化を試み最小ゲノムにおける転写因子数を推定した。

本研究においては幾つかの仮定条件をおいた上で最小ゲノム構築に向けた転写因子数の推定を行った。まず、細胞内におけるmRNA量とタンパク質存在量に関して、先行研究 (de Sousa Abreu, *et al.*, 2009) においても大腸菌において相関が認められており、本研究においても緩やかな相関をタンパク質存在量データと網羅的な大腸菌のマイクロアレイデータの解析により見出した。そして、最も相関の高かったマイクロアレイサンプルGSM101237を用いて回帰式を算出した。そして、この回帰式はマイクロアレイデータをタンパク質存在量へ変換するために用いた。

次に、酵素反応速度論にもとづき反応速度を代謝フラックスと酵素量から求める反応速度式  $V = a[S][E]$  を導出した。このとき、 $a[S]$  は定常状態下において定数的な振る舞いをとると仮定をおいた。この仮定の検証を3種類行った。検証IはGSM101237とは異なるマイクロアレイサンプルを用いた検証である。検証IIは中心代謝に関連する酵素反応に関して詳細な反応速度パラメーターを参照し、参照したモデルにおける $a[S]$ と本研究に用いる $a[S]$ を比較、検証を行った。検証IIIは検証IIにおけるモデルにおける反応速度と代謝流束解析に $a[S]$ を用いてシミュレーションした反応速度の比較、検証をおこなった。そして、これら検証において $a[S]$ は定数的な振る舞いを取ることが示された。特に、検証Iにおける発現量変化に対する代謝流束の安定性は、Ishiiらが定常状態下において行ったマルチオミクスを用いた外的、遺伝的な変化に対して代謝ネットワークが示したロバストネスによって裏付けられる (Ishii, *et al.*, 2007)。

3種類のグループ化の結果、グループに含まれる遺伝子数を連続的に増加させた場合が最も転写因子数を単純化することができた。大腸菌においてORF約4200個をPerez-Ruedaらが予測した転写因子数317個 (Perez-Rueda and Collado-Vides, 2000) で割ると転写因子1個あたり13個の遺伝子を制御していることになり、タイプ1によるグループ化においても野生株のバイオマスを95%保証したときと同程度の制御効率であることが示される。最小ゲノムが遺伝子250個で構成されたとすると、約20個の転写因子を考慮する必要があることが本研究の結果から示唆できる。

## 参考文献

- Arakawa, K., Mori, K., Ikeda, K., Matsuzaki, T., Kobayashi, Y. and Tomita, M. (2003) G-language Genome Analysis Environment: a workbench for nucleotide sequence data mining. *Bioinformatics*, 19, 305–306.
- Arakawa, K., Yamada, Y., Shinoda, K., Nakayama, Y. and Tomita, M. (2006) GEM System: automatic prototyping of cell-wide metabolic pathway models from genomes. *BMC Bioinformatics*, 7, 168.
- Baba, T., Ara, T., Hasegawa, M., Takai, Y., Okumura, Y., Baba, M., Datsenko, K.A., Tomita, M., Wanner, B.L. and Mori, H. (2006) Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection. *Mol Syst Biol*, 2, 2006.0008.
- Barrett, T., Troup, D.B., Wilhite, S.E., Ledoux, P., Evangelista, C., Kim, I.F., Tomashevsky, M., Marshall, K.A., Phillippy, K.H., Sherman, P.M., Muertter, R.N., Holko, M., Ayanbule, O., Yefanov, A. and Soboleva, A. (2011) NCBI GEO: archive for functional genomics data sets—10 years on. *Nucleic Acids Res*, 39, D1005–1010.
- Becker, S.A., Feist, A.M., Mo, M.L., Hannum, G., Palsson, B.Ø. and Herrgard, M.J. (2007) Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox. *Nat Protoc*, 2, 727–738.
- Blank, L.M. and Kuepfer, L. (2010) Metabolic flux distributions: genetic information, computational predictions, and experimental validation. *Appl Microbiol Biot*, 86, 1243–1255.
- Bornstein, B.J., Keating, S.M., Jouraku, A. and Hucka, M. (2008) LibSBML: an API library for SBML. *Bioinformatics*, 24, 880–881.
- Browning, D.F. and Busby, S.J. (2004) The regulation of bacterial transcription initiation. *Nature reviews Microbiology*, 2, 57–65.
- Burgard, A.P., Pharkya, P. and Maranas, C.D. (2003) OptKnock: a bilevel programming framework for identifying gene knockout strategies for microbial strain optimization. *Biotechnol Bioeng*, 84, 647–657.
- Chassagnole, C., Noisommit-Rizzi, N., Schmid, J.W., Mauch, K. and Reuss, M. (2002) Dynamic modeling of the central carbon metabolism of *Escherichia coli*. *Biotechnol Bioeng*, 79, 53–73.
- Colijn, C., Brandes, A., Zucker, J., Lun, D.S., Weiner, B., Farhat, M.R., Cheng, T.Y., Moody, D.B., Murray, M. and Galagan, J.E. (2009) Interpreting expression data with metabolic flux models: predicting *Mycobacterium tuberculosis* mycolic acid production. *PLoS Comput Biol*, 5, e1000489.
- de Sousa Abreu, R., Penalva, L.O., Marcotte, E.M. and Vogel, C. (2009) Global signatures of protein and mRNA expression levels. *Mol Biosyst*, 5, 1512–1526.
- Defresne, P., Hollenberg, H., Husson, B., Tabarki, B., Landrieu, P., Huault, G., Tardieu, M. and Sebire, G. (2003) Acute transverse myelitis in children: clinical course and prognostic factors. *J Child Neurol*, 18, 401–406.
- Delaye, L. and Moya, A. (2010) Evolution of reduced prokaryotic genomes and the minimal cell concept: Variations on a theme. *Bioessays*, 32, 281–287.
- Edwards, J.S. and Palsson, B.Ø. (1998) How will bioinformatics influence metabolic engineering? *Biotechnol Bioeng*, 58, 162–169.
- Edwards, J.S. and Palsson, B.Ø. (1999) Systems properties of the *Haemophilus influenzae* Rd metabolic genotype. *J Biol Chem*, 274, 17410–17416.
- Ellis, T., Adie, T. and Baldwin, G.S. (2011) DNA assembly for synthetic biology: from parts to pathways and beyond. *Integr Biol*, 3, 109–118.
- 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. and Merrick, J.M. (1995) Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. *Science*, 269, 496–512.
- Fong, S.S., Burgard, A.P., Herring, C.D., Knight, E.M., Blattner, F.R., Maranas, C.D. and Palsson, B.Ø. (2005a) In silico design and adaptive evolution of *Escherichia coli* for production of lactic acid. *Biotechnol Bioeng*, 91, 643–648.
- Fong, S.S., Joyce, A.R. and Palsson, B.Ø. (2005b) Parallel adaptive evolution cultures of *Escherichia coli* lead to convergent growth phenotypes with different gene expression states. *Genome research*, 15, 1365–1372.
- Fong, S.S. and Palsson, B.Ø. (2004) Metabolic gene-deletion strains of *Escherichia coli* evolve to computationally predicted growth phenotypes. *Nat Genet*, 36, 1056–1058.
- Forster, A.C. and Church, G.M. (2006) Towards synthesis of a minimal cell. *Mol Syst Biol*, 2, 45.
- Gama-Castro, S., Salgado, H., Peralta-Gil, M., Santos-Zavaleta, A., Muñoz-Rascado, L., Solano-Lira, H., Jimenez-Jacinto, V., Weiss, V., García-Sotelo, J.S., López-Fuentes, A., Porrón-Sotelo, L., Alquicira-Hernández, S., Medina-Rivera, A., Martínez-Flores, I., Alquicira-Hernández, K., Martínez-Adame, R., Bonavides-Martínez, C., Miranda-Ríos, J., Huerta, A.M., Mendoza-Vargas, A., Collado-Torres, L., Taboada, B., Vega-Alvarado, L., Olvera, M., Olvera, L., Grande, R., Morett, E. and Collado-Vides, J. (2011) RegulonDB version 7.0: transcriptional regulation of *Escherichia coli* K-12 integrated within genetic sensory response units (Sensor Units). *Nucleic Acids Res*, 39, D98–105.
- Gibson, D.G., Benders, G.A., Andrews-Pfannkoch, C., Denisova, E.A., Baden-Tillson, H., Zaveri, J., Stockwell, T.B., Brownley, A., Thomas, D.W., Algire, M.A., Merryman, C., Young, L., Noskov, V.N., Glass, J.I., Venter, J.C.C., Hutchison, C.A. and Smith, H.O. (2008) Complete chemical synthesis, assembly, and cloning of a *Mycoplasma genitalium* genome. *Science*, 319, 1215–1220.
- Gil, R., Silva, F.J., Peretó, J. and Moya, A. (2004) Determination of the core of a minimal bacterial gene set. *Microbiol Mol Biol Rev*, 68, 518–537.

- Glass, J.I., Assad-Garcia, N., Alperovich, N., Yooseph, S., Lewis, M.R., Maruf, M., Hutchison, C.A., Smith, H.O. and Venter, J.C.C. (2006) Essential genes of a minimal bacterium. *Proc Natl Acad Sci USA*, 103, 425–430.
- Gosalbes, M.J., Latorre, A., Lamelas, A. and Moya, A. (2010) Genomics of intracellular symbionts in insects. *International journal of medical microbiology : IJMM*, 300, 271–278.
- Hulsen, T., de Vlieg, J. and Alkema, W. (2008) BioVenn - a web application for the comparison and visualization of biological lists using area-proportional Venn diagrams. *BMC Genomics*, 9, 488.
- Ishihama, Y., Schmidt, T., Rappsilber, J., Mann, M., Hartl, F.U., Kerner, M.J. and Frishman, D. (2008) Protein abundance profiling of the *Escherichia coli* cytosol. *BMC genomics*, 9, 102.
- Ishii, N., Nakahigashi, K., Baba, T., Robert, M., Soga, T., Kanai, A., Hirasawa, T., Naba, M., Hirai, K., Hoque, A., Ho, P.Y., Kakazu, Y., Sugawara, K., Igarashi, S., Harada, S., Masuda, T., Sugiyama, N., Togashi, T., Hasegawa, M., Takai, Y., Yugi, K., Arakawa, K., Iwata, N., Toya, Y., Nakayama, Y., Nishioka, T., Shimizu, K., Mori, H. and Tomita, M. (2007) Multiple high-throughput analyses monitor the response of *E. coli* to perturbations. *Science*, 316, 593–597.
- Itaya, M. (1995) An estimation of minimal genome size required for life. *FEBS Lett*, 362, 257–260.
- Itaya, M., Fujita, K., Kuroki, A. and Tsuge, K. (2008) Bottom-up genome assembly using the *Bacillus subtilis* genome vector. *Nat Methods*, 5, 41–43.
- Juhas, M., Eberl, L. and Glass, J.I. (2011) Essence of life: essential genes of minimal genomes. *Trends Cell Biol*, 21, 562–568.
- Keating, S.M., Bornstein, B.J., Finney, A. and Hucka, M. (2006) SBMLToolbox: an SBML toolbox for MATLAB users. *Bioinformatics*, 22, 1275–1277.
- Kim, T.Y., Sohn, S.B., Kim, Y.B., Kim, W.J. and Lee, S.Y. (2011) Recent advances in reconstruction and applications of genome-scale metabolic models. *Curr Opin Biotech*, 23, 1–7.
- Kobayashi, K., Ehrlich, S.D., Albertini, A., Amati, G., Andersen, K.K., Arnaud, M., Asai, K., Ashikaga, S., Aymerich, S., Bessieres, P., Boland, F., Brignell, S.C., Bron, S., Bunai, K., Chapuis, J., Christiansen, L.C., Danchin, A., Debarbouille, M., Dervyn, E., Deuerling, E., Devine, K., Devine, S.K., Dreesen, O., Errington, J., Fillinger, S., Foster, S.J., Fujita, Y., Galizzi, A., Gardan, R., Eschevins, C., Fukushima, T., Haga, K., Harwood, C.R., Hecker, M., Hosoya, D., Hullo, M.F., Kakeshita, H., Karamata, D., Kasahara, Y., Kawamura, F., Koga, K., Koski, P., Kuwana, R., Imamura, D., Ishimaru, M., Ishikawa, S., Ishio, I., Coq, D.L., Masson, A., Mauriel, C., Meima, R., Mellado, R.P., Moir, A., Moriya, S., Nagakawa, E., Nanamiya, H., Nakai, S., Nygaard, P., Ogura, M., Ohanan, T., O'reilly, M., O'rourke, M., Pragai, Z., Pooley, H.M., Rapoport, G., Rawlins, J.P., Rivas, L.A., Rivolta, C., Sadaie, A., Sadaie, Y., Sarvas, M., Sato, T., Saxild, H.H., Scanlan, E., Schumann, W., Seegers, J.F.M.L., Sekiguchi, J., Sekowska, A., Seror, S.J., Simon, M., Stragier, P., Studer, R., Takamatsu, H., Tanaka, T., Takeuchi, M., Thomaides, H.B., Vagner, V., van Dijl, J.M., Watabe, K., Wipat, A., Yamamoto, H., Yamamoto, M., Yamamoto, Y., Yamane, K., Yata, K., Yoshida, K., Yoshikawa, H., Zuber, U. and Ogasawara, N. (2003) Essential *Bacillus subtilis* genes. *Proc Natl Acad Sci USA*, 100, 4678–4683.
- Koonin, E.V. (2000) How many genes can make a cell: the minimal-gene-set concept. *Annu Rev Genom Hum G*, 1, 99–116.
- Liu, M., Durfee, T., Cabrera, J.E., Zhao, K., Jin, D.J. and Blattner, F.R. (2005) Global transcriptional programs reveal a carbon source foraging strategy by *Escherichia coli*. *J Biol Chem*, 280, 15921–15927.
- Lu, P., Vogel, C., Wang, R., Yao, X. and Marcotte, E.M. (2007) Absolute protein expression profiling estimates the relative contributions of transcriptional and translational regulation. *Nat Biotechnol*, 25, 117–124.
- McCutcheon, J.P., McDonald, B.R. and Moran, N.A. (2009) Origin of an alternative genetic code in the extremely small and GC-rich genome of a bacterial symbiont. *PLoS Genet*, 5, e1000565.
- McCutcheon, J.P. and von Dohlen, C.D. (2011) An interdependent metabolic patchwork in the nested symbiosis of mealybugs. *Curr Biol*, 21, 1366–1372.
- Meadows, A.L., Karnik, R., Lam, H., Forestell, S. and Snedecor, B. (2010) Application of dynamic flux balance analysis to an industrial *Escherichia coli* fermentation. *Metab Eng*, 12, 150–160.
- Mizoguchi, H., Sawano, Y., Ichi Kato, J. and Mori, H. (2008) Superpositioning of deletions promotes growth of *Escherichia coli* with a reduced genome. *DNA Res*, 15, 277–284.
- Moya, A., Gil, R., Latorre, A., Peretó, J., Garcillán-Barcia, M.P. and de la Cruz, F. (2009) Toward minimal bacterial cells: evolution vs. design. *FEMS Microbiol Rev*, 33, 225–235.
- Mulkidjanian, A.Y., Koonin, E.V., Makarova, K.S., Mekhedov, S.L., Sorokin, A., Wolf, Y.I., Dufresne, A., Partensky, F., Burd, H., Kaznadzey, D., Haselkorn, R. and Galperin, M.Y. (2006) The cyanobacterial genome core and the origin of photosynthesis. *Proc Natl Acad Sci USA*, 103, 13126–13131.
- Mushegian, A.R. and Koonin, E.V. (1996) A minimal gene set for cellular life derived by comparison of complete bacterial genomes. *Proc Natl Acad Sci USA*, 93, 10268–10273.
- Nakabachi, A., Yamashita, A., Toh, H., Ishikawa, H., Dunbar, H.E., Moran, N.A. and Hattori, M. (2006) The 160-kilobase genome of the bacterial endosymbiont *Carsonella*. *Science*, 314, 267.
- Nie, L., Wu, G. and Zhang, W. (2006) Correlation between mRNA and protein abundance in *Desulfovibrio vulgaris*: a multiple regression to identify sources of variations. *Biochem Biophys Res Commun*, 339, 603–610.
- O'Malley, M.A., Powell, A., Davies, J.F. and Calvert, J. (2008) Knowledge-making distinctions in synthetic biology. *Bioessays*, 30, 57–65.
- Orth, J.D., Thiele, I. and Palsson, B.Ø. (2010) What is flux balance analysis? *Nat Biotechnol*, 28, 245–248.

- Pagani, I., Liolios, K., Jansson, J., Chen, I.M.A., Smirnova, T., Nosrat, B., Markowitz, V.M. and Kyrpides, N.C. (2012) The Genomes OnLine Database (GOLD) v.4: status of genomic and metagenomic projects and their associated metadata. *Nucleic Acids Res*, 40, D571–579.
- Pál, C., Papp, B., Lercher, M.J., Csermely, P., Oliver, S.G. and Hurst, L.D. (2006) Chance and necessity in the evolution of minimal metabolic networks. *Nature*, 440, 667–670.
- Pérez-Rueda, E. and Collado-Vides, J. (2000) The repertoire of DNA-binding transcriptional regulators in *Escherichia coli* K-12. *Nucleic Acids Res*, 28, 1838–1847.
- Purnick, P.E.M. and Weiss, R. (2009) The second wave of synthetic biology: from modules to systems. *Nat Rev Mol Cell Bio*, 10, 410–422.
- Ramakrishnan, S.R., Vogel, C., Prince, J.T., Li, Z., Penalva, L.O., Myers, M., Marcotte, E.M., Miranker, D.P. and Wang, R. (2009) Integrating shotgun proteomics and mRNA expression data to improve protein identification. *Bioinformatics*, 25, 1397–1403.
- Raman, K. and Chandra, N. (2009) Flux balance analysis of biological systems: applications and challenges. *Brief Bioinformatics*, 10, 435–449.
- Reaves, M.L. and Rabinowitz, J.D. (2010) Metabolomics in systems microbiology. *Curr Opin Biotech*, 22, 1–9.
- Reed, J.L., Vo, T.D., Schilling, C.H. and Palsson, B.Ø. (2003) An expanded genome-scale model of *Escherichia coli* K-12 (iJR904 GSM/GPR). *Genome Biol*, 4, R54.
- Scheer, M., Grote, A., Chang, A., Schomburg, I., Mounetto, C., Rother, M., Schöngren, C., Stelzer, M., Thiele, J. and Schomburg, D. (2011) BRENDA, the enzyme information system in 2011. *Nucleic Acids Res*, 39, D670–676.
- Schellenberger, J., Park, J.O., Conrad, T.M. and Palsson, B.Ø. (2010) BiGG: a Biochemical Genetic and Genomic knowledgebase of large scale metabolic reconstructions. *BMC Bioinformatics*, 11, 213.
- Shinfuku, Y., Sorpitorn, N., Sono, M., Furusawa, C., Hirasawa, T. and Shimizu, H. (2009) Development and experimental verification of a genome-scale metabolic model for *Corynebacterium glutamicum*. *Microb Cell Fact*, 8, 43.
- Smolke, C.D. (2009) Building outside of the box: iGEM and the BioBricks Foundation. *Nat Biotechnol*, 27, 1099–1102.
- Szostak, J.W., Bartel, D.P. and Luisi, P.L. (2001) Synthesizing life. *Nature*, 409, 387–390.
- Thiele, I. and Palsson, B.Ø. (2010) A protocol for generating a high-quality genome-scale metabolic reconstruction. *Nat Protoc*, 5, 93–121.
- Tian, Q., Stepaniants, S.B., Mao, M., Weng, L., Feetham, M.C., Doyle, M.J., Yi, E.C., Dai, H., Thorsson, V., Eng, J., Goodlett, D., Berger, J.P., Gunter, B., Linseley, P.S., Stoughton, R.B., Aebersold, R., Collins, S.J., Hanlon, W.A. and Hood, L.E. (2004) Integrated genomic and proteomic analyses of gene expression in Mammalian cells. *Mol Cell Proteomics*, 3, 960–969.
- Wittig, U., Kania, R., Golebiewski, M., Rey, M., Shi, L., Jong, L., Alga, E., Weidemann, A., Sauer-Danzwirth, H., Mir, S., Krebs, O., Bittkowski, M., Wetsch, E., Rojas, I. and Müller, W. (2012) SABIO-RK—database for biochemical reaction kinetics. *Nucleic Acids Res*, 40, D790–796.
- Yizhak, K., Tuller, T., Papp, B. and Ruppin, E. (2011) Metabolic modeling of endosymbiont genome reduction on a temporal scale. *Mol Syst Biol*, 7, 479.