

Network motif による遺伝子発現調節

東京大学総合文化研究科 石原 秀至*

1 はじめに：背景

実験技術の進歩により、細胞内で働く遺伝子、蛋白、代謝物などの分子的データが蓄積されてきた。それに伴いデータを解釈するための情報技術やモデリングの必要性も増大している。実際、多くの生物的過程/機能は種々多様の複雑な分子間の相互作用を通して現れるので、与えられた分子レベルのデータからそれらが果たす機能を類推するのは基礎的かつ重要な問題である。例えば発生においては相互作用しあう遺伝子の一連の発現を通じて形態形成が行われる。このような問題に対して、生物の機能を理解するための一般的な機構 (general principle) を期待することはできるのだろうか？ 近年分子間の相互作用からなるネットワーク構造について、small-world 性や scale-free 性など生物種を越えて普遍的に見られる統計的な性質がいくつか見出されている。これらの性質の生物における役割や意義については十分理解されているわけではないが、上の問題に対しての手がかりを与えていると考えられる。

U. Alon らは幾種類かの部分構造 (network motif) が生物のネットワークにおいて有意に多く含まれていることを見出した [1]。統計的有意さは、各要素の入出力の矢印の数が実際の生物ネットワークのものと同じランダムなネットワークを多数生成し、その平均との比較による。この比較の結果、実際の生物ネットワークに有意に多く見出される部分構造を network motif と呼ぶ。その中でも Feed-forward loop (FFL) と呼ばれる 3 つの要素からなる network motif (図 1) は、大腸菌・酵母の遺伝子転写ネットワークや線虫の神経ネットワークに共通に現れる典型的なものである。例えば大腸菌の転写ネットワークでは 42 個の FFL があるが、ランダムなネットワークでは平均 8.1 個でしかない (表 1)。Network motif はその多さや種を越えた共通性から基本的な機能単位をしていると考えられる。

転写制御の場合、FFL は図 1 のような 3 つの遺伝子 (オペロン) からなるネットワークであり、X が Y, Z を、Y が Z の活性を制御している。制御は正 (active)、または負 (repressive) であるとし、X から Z への二つの制御経路がどちらも正 (または負) に働く場合を coherent FFL、両者で異なる場合を incoherent FFL と言う。一つの FFL の機能については、X の活性を入力、Z の活性を出力として、理論的・実験的に調べられた [3]。各制御の正負やパラメータに応じて幾つかの機能が報告されている。[persistent detection] 一過性の入力には応答せず Z の活性には継続入力が必要; [pulse generator] 入力変化に対して一過性の活性化のみを示す; [acceleration of response] 単一の 2 体制御より応答が早い。各々は一つの FFL の機能であるが、実際の生物過程で、あるいはネットワークの中でどのような役割を果たしているかは明らかではない。特にもともとの動機であった、期待すべき “general principle” との関連は議論されつくされていない。

本講演のもう一つの主題は発生である。ショウジョウバエ (*Drosophila*) は多細胞生物ではその転写制御が最もよく理解されているモデル生物であり、我々は、Network motif がショウジョウバエ (*Drosophila melanogaster*) の発生において本質的な役割を果たしていることを示す。勿論、発生は最も興味深い生物過程の一つであり、複雑な分子間の相互作用を通して受精卵から高度な体制が形成される。特にその初期における重要な過程として体節形成がある。体節形成はいくつかの遺伝子の発現によって行われる。このような遺伝子発現の空間パターンはどのように行われるのであろうか？ Wolpert は morphogen 因子が空間勾配を形成し、その濃度に応じて胚中の各場所で遺伝

*shuji@complex.c.u-tokyo.ac.jp

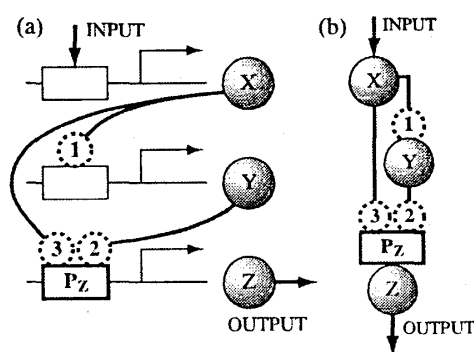


図 1: (a) Feed-forward loop network motif. X が Y、Z の転写因子をコードし、Y が Z の転写因子をコードする。1, 2, 3 それぞれの制御は活性か抑制である。X、Y は Z の *cis* 領域 (P_Z) を介して制御される。本講演では AND 的か OR 的に働くかと仮定する。(b) FFL を回路的に表したものの。

子が発現する (morphogen が位置情報を担う) というモデルを提唱した [4]。 *Drosophila* の発生では幾つかの母親由来 (maternal) 因子、特に *bicoid* (*bcd*) が初期胚において勾配を形成し、それが胚に極性をもたらし、かつ、実際 Wolpert 説の morphogen として働いていることがわかってきた[†][5]。さて、Wolpert 理論は位置情報の起源を説明はしたが、位置情報に基づいた体節化などのパターン形成メカニズムまでは明らかにしていない。例えば *Drosophila* 胚における遺伝子 *even-skipped* (*eve*) は 7 本縞様に発現し、体節化に重要な役割を果たす。そのような *eve* の縞の形成については、種々の mutant 実験により、それは *bcd* を含む各遺伝子の複雑な相互作用を通して起こり、また、実験を通して発生における遺伝子の階層的分類とそのカスケード的な制御描像が確立した [6]。その上で、このような複雑な遺伝子間制御ネットワークからパターンを形成する (*Drosophila* だけに限らない) 一般的メカニズムを明らかにすることが重要であろう。理論的には *bcd* 勾配と *eve* を直接制御する gap 遺伝子群の制御ネットワークを閾値型制御と空間拡散を取り入れたモデル化が多くなされている [7]。これらの研究では多数のパラメータを simulated annealing 等の手法で実際の発生にあうように決め、結果得られたモデルは発生を再現する。しかしながらパターン発生のメカニズムの「理解」にはまだ至っていない。

以上の背景をふまえ、我々は次節から FFL のある機能が発生のパターン形成に対して重要な役割を果たしていることを示す。

2 大腸菌と *Drosophila* の転写ネットワーク

はじめに、転写ネットワークの構造を特徴付ける[‡]。ここでは大腸菌 (*E. Coli*) と *Drosophila* 初期発生ものを比較する。データとしては、*E. Coli* については Shenn-Orr らがリリースしたもの [8][§]、*Drosophila* については GeNet [9] [¶] を用いる。両者とも文献に基づいたデータベースである。他のデータベースを用いても、下に述べる性質は変わらなかった。

まず、両者で FFL がどの程度多いのかを見るために、前に述べた意味でのランダムネットワークと実際のネットワークを比較する。大腸菌、*Drosophila* とともに FFL を多く含むことがわかる (表 1)。大腸菌では incoherent FFL が少ないのに対し (6/36)、*Drosophila* では約半数である (43/83)。

次に FFL 同士がどの様につながっているのかを調べる。もとのネットワークから FFL だけを取

[†]他の生物種でもある因子 (morphogen) の濃度に応じて分化経路を変えることは多く観察される。例えばアフリカツメガエルにおいてはアクチビン濃度に応じて細胞運命が変わる。

[‡]内容は一部報告済みのものと重なるがいちいち指摘しない

[§]<http://www.weizmann.ac.il/mcb/UriAlon>

[¶]<http://www.csa.ru/Inst/gorb.dep/inbios/genet/genet.htm>

Database	genes	regulations	FFL	iFFL
<i>E. Coli</i>	423	578	42 (8.1 ± 3.2)	6 (/36)
<i>Drosophila M.</i>	122	309	157 (103.7 ± 11.1)	43 (/83)

表 1: 大腸菌 (*E. Coli*)、ショウジョウバエ (*Drosophila*) の二つのデータベースに対する遺伝子 (*E. Coli* に対してはオペロン)、制御、FFL、incoherent FFL の数。FFL の項の括弧はランダムネットワークの平均 FFL 数とその標準偏差。Incoherent FFL を数えるときには活性・抑制がはっきりしていないものを除いた。括弧の中はその場合の FFL の数である。

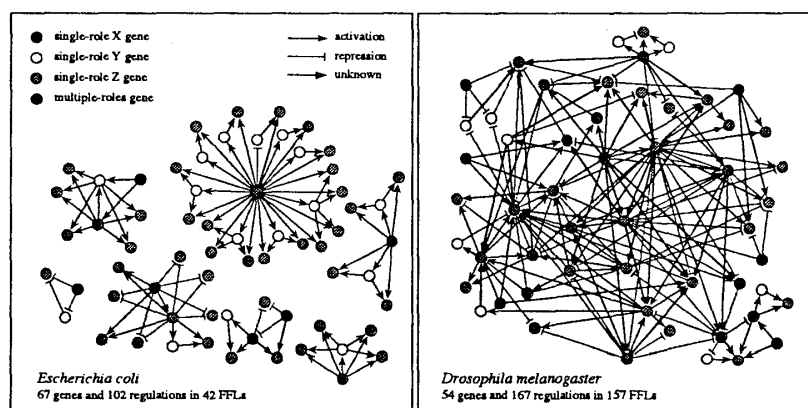


図 2: FFL のネットワーク (a) *E. Coli*, (b) *Drosophila*。矢印は活性、抑制、濃度依存の制御を表す。

り出したものが図 2 である。まず、ほとんどの FFL は独立していないことがわかる (isolated FFL は大腸菌で一つのみ)。また、一見して大腸菌、*Drosophila* で異なる構造が見られる。もう少し詳しく見ると、(i) 大腸菌では 67 の要素で 42 の FFL を作っている。一方 *Drosophila* では 54 の要素で 157 の FFL を作っている。即ち *Drosophila* では大腸菌と比べ一つの要素が何度も異なる FFL に現れる。二つの FFL が同じ要素あるいは制御を共有していることを FFL の相互作用と言おう。上の結果は *Drosophila* においては FFL からなるネットワークはよりこみいっていることを表す。また、大腸菌においては FFL ネットワークはいくつかのブロックに分かれている。(ii) 相互作用にはいくつかの種類がある。相互作用している二つの FFL に共通な要素が、その二つの FFL に対して異なる役割 (図 1 の X,Y,Z) をしている時、multi-roles gene と呼ぼう。(例えば図 3(b) の Z_1 は図 3 の Z と X の役割をしている。) その様な multi-roles gene を含む相互作用の割合を測ると、大腸菌では約 2% しかないが、*Drosophila* では約 18% ある。これは (図でもわかるように) 大腸菌では X,Y を共通にもつ一入力多出力の並列的な相互作用が多く、短い経路が多いのに対し、*Drosophila* では比較的長くカスケード的・直列的な経路を形成していることを反映している^{||}。

この様な構造の違いは二つのネットワークが果たす役割の違いを反映していると考えられる。特に *Drosophila* のネットワークでは FFL がよりこみいった相互作用をしており、おそらく一つの FFL の機能ではなく相互作用によって現れる機能が重要であろう。次節で我々が見出した FFL の新しい機能を説明する。

^{||} 最近、大腸菌の二つのデータベースを組み合わせた結果大腸菌の転写ネットワークでも FFL が [8] で見出されたものほど単純ではないという報告がされた [14]。

3 Concentration detectability of FFL networks

この節では FFL の concentration detector としての機能を説明する。single FFL の機能をもとに、FFL が相互作用することで出来る複雑な応答を調べる。

incoherent FFL as a concentration detector : 図 1 で gene X の発現量を入力、gene Z の発現量を出力として、Z の X に対する応答を調べる^{**}。ここで各制御に対しては簡単な Hill の式を仮定する。即ち、活性化に対しては $f(X; K, H) = X^H / (X^H + K^H)$ 、抑制に対しては $f(X; K, H) = K^H / (X^H + K^H)$ とする。これは閾値を K 付近に持つ sigmoid 的な反応式となる。Z の制御に関しては X、Y 二つの regulator が関与するので、ここでは AND 的に働くとしよう。この時 coherent な FFL では単調な関数になるのに対して、incoherent な FFL では、あるパラメタ条件で $[Z]$ が $[X]$ に対して一山型になる (図 3(a))。Z が X の発現量がある範囲にある時にのみ発現するので、この回路は入力 X の concentration detector となっている。この範囲は X から Z への二つの経路のそれぞれの閾値に対応している (図 3(a2))。

Responses of interactiong FFL motifs : 一つの incoherent FFL では一山の応答を示すことを見たが、この性質により、FFL が相互作用すると複雑な応答を示すことになる。その例としてここでは直列 FFL (図 3(b)) と並列 FFL (図 3(c)) を見てみよう。直列 FFL では初めの FFL の出力が次の FFL の入力になっている (Z_1 が multi-role)。この様な回路では $[Z_1]$ が $[X]$ に対し一山の応答を示し、 $[Z_2]$ が $[Z_1]$ に対し一山の応答を示すので、結果として $[Z_2]$ は $[X]$ に対し二山型の応答を示すことになる (図 3(b))。したがって直列 FFL では $[X]$ の二つの発現量域で Z_2 の活性化が起こる。この過程は doubling であり、直列を増やすと 4 領域で活性化が起こる。図 3(c) のような並列 FFL でも多重の $[X]$ 領域で Z の活性化が起こり得る。この回路においては X から Z への 3 本の制御経路があり、それぞれの閾値が Z の活性領域を決定するのは、単一 FFL の場合と同じであり、図では左から K_{Y_2}, K_{Y_1}, K_{XZ} (それぞれ対応する経路の閾値) である。このように FFL の相互作用によって出力遺伝子の活性は $[X]$ の複数の活性領域で発現することになる。

4 Pattern formation by FFL network

前節までの FFL の concentration detectability によって空間的なパターンを作ることが出来ることを示す。前述したように生物の発生過程ではしばしば胚中に morphogen 勾配がある。例えば *Drosophila* では、母親由来の *bcd*-mRNA が胚頭部に固定されており、局所的な翻訳と胚内への拡散^{††}により Bcd-タンパクは頭部から尾部にかけて指数的に減衰する勾配を形成する。この様な分子が FFL 中において X の転写因子として働く場合を考えると、Z が胚のある空間領域でのみ発現することは容易に理解できよう。図 4(a) に示すようにそれぞれの経路の閾値に対応する空間範囲で Z が活性化される。より複雑なパターンは FFL の相互作用を通して現れる。直列 FFL の入力が morphogen として働けば、その出力遺伝子はストライプ等の空間パターンを形成して活性化する。図 4(b) の例では Z_3 が 4 本縞を形成して発現する。

5 Morphogenesis of *Drosophila M.*

我々は *Drosophila* の転写ネットワークでは FFL が多く、かつ複雑な相互作用をしていることを見た。これは前節で述べたメカニズムが実際に働いていることを示唆する。ここでは胚中の *eve* の発現を構成し、そのネットワークを解析する。GeNet[9] や他の幾つかのデータベースを参考に Network を構成した。^{††} 図 5(a) にその構成した転写ネットワークを示す。Feed-back や mutual repression

^{**}定常状態だけを調べる。即ち、入力濃度 $[X]$ に応じた $[Z]$ の値を見る。ここで $[X]$ は X の発現量。

^{††}*Drosophila* は初期ステージでは一細胞多核。

^{††}ここで用いたデータ、また、cis-制御やパラメタについての詳細は [11] を参照のこと。*eve1 stripe* についてはその regulator が判っていないのでここでは無視した。

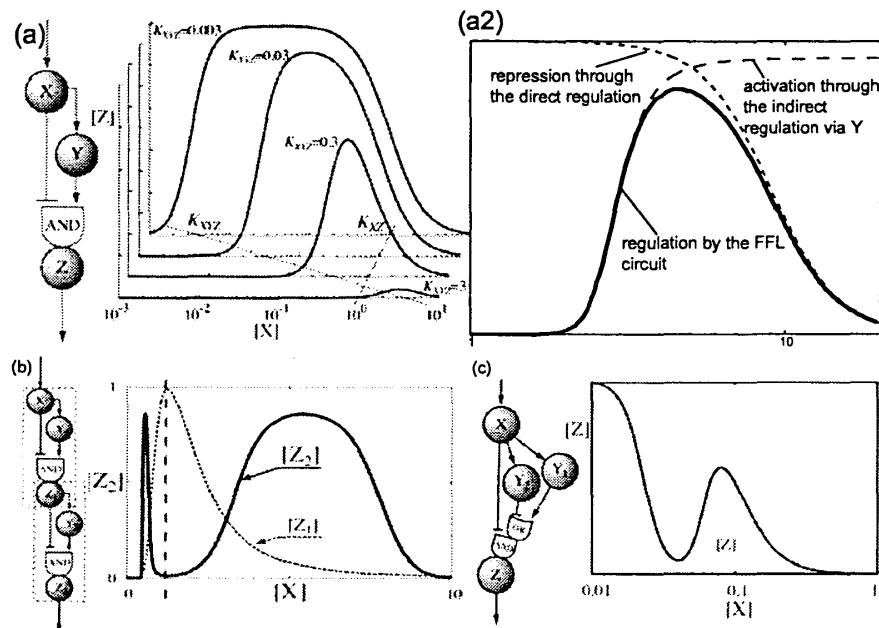


図 3: (a) incoherent FFL の一山型応答。左図のような incoherent FFL に対して、Z の発現量 $[Z]$ を X の発現量 $[X]$ の関数として表したもの。($K_{XZ} = 1$, $K_{XYZ} = 0.003, 0.03, 0.3$, and 3 .) (a2) 一山型応答において Z の発現領域は二つの経路のそれぞれの閾値で決まる。(b) 直列 FFL とその応答。(c) 並列 FFL とその応答。

をとりいれていない。このネットワークにおいては Bcd, Nos の勾配が用意されており、それがネットワークへの入力となっている。出力としての *eve* 遺伝子の発現パターンを図 5(a) に示す。この図 5(a) のネットワークにおいては *Cad* を除いて全ての因子が incoherent FFL のメンバーになっている。Incoherent FFL による Morphogen 空間勾配の concentration detection が働いている。

6 議論・まとめ・展望

我々は本講演で次のことを見た。(i) FFL の concentration detectability によって形態形成が可能であること。(ii) *Drosophila M.* の体節構造形成には実際にその機構が働いていること。以下この機構について議論する。

Turing Pattern との比較： 形態形成の理論は A. Turing によって始められた [12]。活性-抑制因子の二成分の反応拡散系において一様状態の不安定化がパターンを発生させる。脊椎動物の発生や体表パターンにはこれに近い機構が働いていると考えられている [13]。我々が議論した FFL による機構と比較すると、最も大きな違いは、縞の数のコーディングの仕方であろう。(仮想的に) 胚のサイズを大きくした時、Turing Pattern では縞の数が多くなってしまいが、FFL による機構では数は変わらない。ただし Morphogen の空間勾配が指数的である場合はプロポーションは保たれない。
Robustness： パラメタの変化に対して表現型がどれくらい影響を受けると言うロバストネスが良く議論される。FFL による機構では、各遺伝子発現の場所には閾値の大きさが重要なので、その大きさの順序が変わらない限り基本的な体節構造は変わらない。また、mutual regulation や feed-back が働いているので、それによる調節も重要だろう。

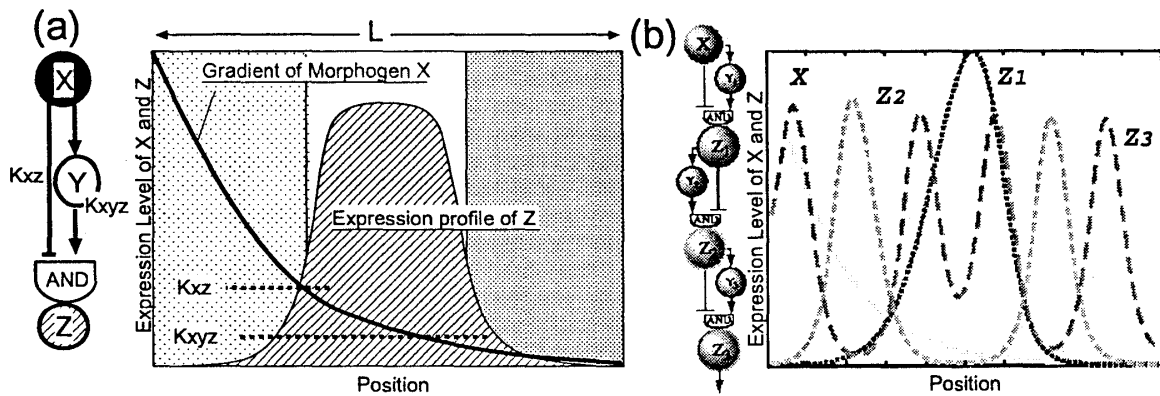


図 4: Incoherent FFL のよって空間パターンを作る仕組み。(a) 空間勾配をもつ morphogen が iFFL ネットワークの X gene として働いているならば、Z gene は空間のある領域で発現する。(b) 直列的な FFL による doubling で縞様の空間パターンができる。

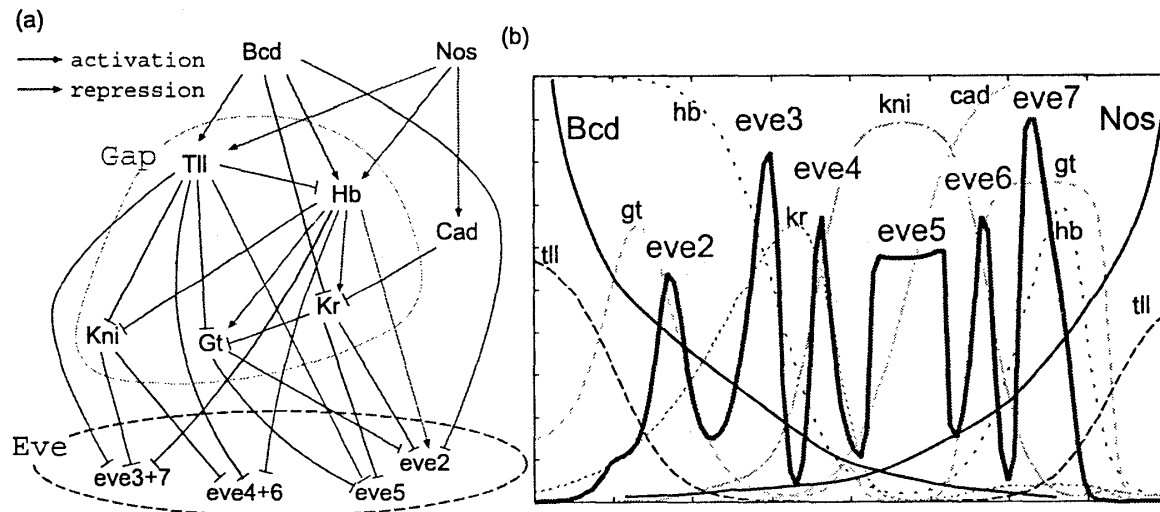


図 5: (a) *Drosophila* 発生初期の転写ネットワークモデル。(b) (a) のネットワークによる空間発現。

7 おわりに

本講演の内容は藤本仰一氏 (東大総文)、柴田達夫氏 (広大理) との共同研究に基づく。深く感謝する。また、より詳しい内容を知りたい方は [10, 11] を参照していただきたい。

- [1] R. Milo, S. S. Shen-Orr, S. Itzkovitz et al. *Science* **298**, 824-827. (2002)
- [2] S. S. Shen-Orr, R. Milo, S. Mangan, & U. Alon *Nat. Genet.* **31**, 64-68. (2002)
- [3] S. Mangan & U. Alon *Proc. Natl. Acad. Sci. USA* **100**, 11980-11985. (2003)
- [4] L. Wolpert *J. Theor. Biol.* **25**, 1-47. (1969)
- [5] S. F. Gilbert "Developmental Biology" Massachusetts: Sinauer (2000)
- [6] C. Nüsslein-Volhard & E. Wieschaus *Nature* **287** 795-801 (1980)
- [7] J. Jaeger, S. Surkova, M. Blagov et al. *Nature* **430**, 368-371 (2004)
- [8] S. S. Shen-Orr, R. Milo, S. Mangan, & U. Alon, *Nat. Genet.* **31**, 64-68. (2002)
- [9] V. N. Selov, A. V. Spirov & M. G. Samsonova *Bioinformatics* **14**, 546-547. (1998)
- [10] S. Ishihara, K. Fujimoto, & T. Shibata *submitted*
- [11] S. Ishihara Ph. D thesis, Chap. 2-3.
- [12] A. M. Turing *Philos. Trans. Roy. Soc. London. B* **237**, 37-72. (1952)
- [13] S. Kondo & R. Asai *Nature* **376**, 765-768. (1995)
- [14] H.-W. Ma, B. Kumar, U. Ditges, et al. *Nucl. Acids Res.* **32** 6643 (2004)