転写因子における DNA- 蛋白質相互作用

奈良先端科学技術大学院大学バイオサイエンス研究科 箱嶋敏雄

Toshio HAKOSHIMA: DNA-protein Interactions by Transcription Factors

The diversity of transcription factor-DNA complexes reviewed in this paper marks the extent of discovery since the first transcription factor-DNA co-crystal structures appeared less than ten years ago. The DNA-binding domains of transcription factors were classified into families, many of which have been identified by comparative amino acid sequence analysis. The three-dimensional views of their respective families revealed new designs as well as unanticipated similarities to older ones. Several hoped-for simplicities in their DNA-recognition mechanism have proved illusory, but other unexpected regularities have emerged.

1. 転写は蛋白質によって制御されている

転写には、DNA 依存的に RNA を合成する酵素である RNA ポリメラーゼ以外に、転写因子、基本転写因子あるいは介在因子と呼ばれる蛋白質群が関与する (図1)。転写因子は遺伝子発現のスイッチ役であり、転写を抑制するリプレッサーと活性化するアクチベータに分類される。原核細胞ではリプレッサーによる負の制御が主流であるが、酵母や植物あるいは動物など核をもつ真核細胞ではアクチベータによる正の制御が主流である。このことは、細菌などの原核細胞では DNA が常にさまざまな蛋白質の結合が可能な状態にあるために転写開始が比較的容易であるのに対して、核をもつ真核細胞では、遺伝子 DNA はヒストン蛋白質などによってヌクレオソーム (DNA がヒストンにより折り畳まれた染色体の基本構造体) と呼ばれる高次構造体を形成することにより他の蛋白質との接触が大きく制限されているので、これらの構造体の解消などのプロセスが必要であることと関係して

いると考えられる。実際、ヌクレオソーム構造を解消する転写因子も見つかっている。また、原核細胞では通常 1 種類 (多くとも 2 種類) の転写因子が一つの転写単位 (レギュロン) を制御しているのに対して、真核細胞では複数の転写因子が転写開始に必要とされる。基本転写因子は、転写開始位置の認識や複数の転写因子からの情報を集約する役割が考えられている。特に、基本転写因子の多い真核細胞では、ヌクレオソーム構造の解除を担う蛋白質群も含む。介在因子は DNA と直接には相互作用しないが、転写因子と基本転写因子との蛋白質-蛋白質相互作用を仲介する。現在、真核細胞で数十の介在因子蛋白質が見つかっている。

転写に関わる蛋白質の三次元構造決定は、1980 年代の初頭から始まり、後半には化学合成された DNA オリゴマーとの複合体の構造決定が報告されるようになった。1997 年 1 月の PDB (Protein Data Bank, Brookhaven, USA) では 217 件の蛋白質-核酸複合体のエントリーがあり、現在 50 種類あまりの蛋白質/DNA 複合体の構造が決定されている。

2. 転写因子は複数の分子機能をもつ

転写因子の基本機能は DNA の特定配列部位への結合と転写活性の阻害あるいは活性化であるが、ホルモンなどのシグナル分子の受容とその応答や、他の転写因子との会合を通じた転写制御ネットワークの形成などの機能をもつ (図2)。これらの機能は、物理的には構造ドメインと呼ばれる三次元構造を保つことのできるペプチド鎖

図1 転写開始の概念図 (Initiation of transcription)

図2 転写因子の機能ドメインの概念図 (Transcription factor has three domains)

の単位 (“かたまり”) に対応しており、機能ドメインと呼ばれる。切り出してきた機能ドメインはそれ自体で機能する。転写因子は複数の機能ドメインが連なった比較的柔軟性に富んだ分子である。また、溶液中で会合しやすく、多分散系になりやすい。特に、真核細胞由来の転写因子は大腸菌を用いた高発現系が働かない場合が多い。これらの諸問題が全長蛋白質 (十数 kDa ~ 数十 kDa 以上) としての転写因子の構造研究を困難なものとしている。これまでの構造研究の主流は DNA の塩基配列を認識する DNA 結合ドメインである。一般に、このドメインにはアルギニン、リジンなどの塩基性のアミノ酸残基に富んでおり、酸性である核酸との相互作用が有利になるように設計されている。多くの転写因子はカルボキシル末端あるいはアミノ末端に DNA 結合ドメインをもつ。

3. DNA 結合ドメインは特徴的な構造モチーフをもつ

二重らせん DNA の塩基対と糖-リン酸骨格は、それぞれ、らせんの内側と外側に位置する。糖-リン酸骨格の間には2つの溝が形成され、塩基対の一部が露出している。広いほうの溝を主溝 (major groove)、狭い方の溝を副溝 (minor groove) と呼ぶ (図3 (a) (b))。DNA-蛋白質複合体結晶中では、DNA オリゴマーはらせん軸がとぎれなく繋がるように積み重なって、まるで一本の長い鎖のようにパッキングしている (図3 (c))。DNA の塩基配列を “読みとる” DNA 結合ドメインの大きさは、5kDa ~ 15kDa 程度である。ほとんどの DNA 結合ドメインは主溝側からの塩基との直接相互作用を通して塩基配列を認識するが、少数のものは副溝からも塩基配列を認識する。また、溝にはまり込んで塩基と直接相互作用する部位が、 α -らせんか β -シートあるいは二次構造をとらないループであるかによって3群に大別される。さらに、DNA の塩基配列を直接に認識している部位周辺の三次元構造上の特徴 (これを DNA 結合モチーフと呼ぶ) から、比較的

少数の類型に分類できる。現在のところ、十種類あまりの DNA 結合モチーフが知られており (表1)、アミノ酸配列の相同性が高いものと低いもの、また、認識配列も類似性が高いものや変化の富んだものがある。一般に、多くの転写因子は二量体を形成して回文状の塩基配列をもつ DNA に結合する。回文状 (palindromic) の塩基配列とは、Watson-Crick の塩基対を形成する相補的な二本の DNA 鎖の配列が同じになる配列のことで、例えば後述する HLH モチーフの認識する E-box と呼ばれる6塩基対からなる塩基配列は、

のように完全な、あるいは部分的な回文状配列 (下線部) である。また、単量体で対称性のない塩基配列を認識するものもある。これらの特徴を念頭に置いて、実際に、DNA との複合体の三次元構造が決定されて、DNA の塩基配列の認識まで実験的に確認された例を引用しつつ以下の項で各々のモチーフを紹介する。ほとんどの DNA-蛋白質複合体の構造は、X 線結晶解析で決定されている。

4. HTH と wHTH モチーフは細菌からヒトまで広く分布する

α -らせんにより塩基配列を認識する代表的な DNA 結合モチーフには、2本の α -らせんが短いターン構造で結ばれた HTH (helix-turn-helix) モチーフがあり、原核細胞から真核細胞まで広く分布している。特徴としては、アミノ末端側にもう一本の α -らせんをもち、合計3本の α -らせんが疎水的なコアを形成することによりドメイン構造を保っている。カルボキシル末端側の α -らせんが二重らせん DNA の主溝に入り、塩基配列を認識する (図4 (a))。フィードバック阻害による代謝系酵素等の遺伝子の発現を制御している細菌由来のリプレッサー蛋白質のほとんどがこのモチーフをもち、二量体を形成して回文状の塩基配列に結合する。主溝に入り込む各サブユニットの α -らせんは、およそ 34Å 離れている場合が多い。これは DNA 二重らせんの1巻き分のピッチに相当しており、 α -らせんは二重らせん構造の同じ側面から主溝に入り込むことになる (図4 (b))。塩基対表面はらせんの曲面を形成しているために、各 α -らせんによって直接に認識される塩基配列は数塩基対 (通常5塩基対以内) であるが、二量体で結合するために、糖やリン酸部分との相互作用も含めると、15 から 25 塩基対が蛋白質と相互作用する。一方、環境応答系の遺伝子の発現制御をしている転写因子、例えば細胞の浸透圧応答系を制御している OmpR なども HTH モチーフあるいは wHTH モチーフ (後述) をもつ。これらの認識配列は必ずしも対称をもたず、二量体として結合するのか、あるいは単量体や多量体として結合するのかは明らかではない。これらは遺伝子発

図3 二重らせんDNA (Double helical structure of DNA)

(a) 二重らせんDNAの2つの溝 (groove) . 広い主溝 (major groove) と狭い副溝 (minor groove) があり, 主溝では塩基の縁が大きく露出している. 左の図は標準的なB型の二重らせん構造, 右は酵母の転写因子PHO4に結合した17塩基対のDNAオリゴマーの複合体結晶¹⁾中での二重らせん構造, 中央はそれらを重ね合わせた図である.

(b) DNA塩基対とらせんの溝との関係. アデニン (A) ・チミン (T) 塩基対とグアニン (G) ・シトシン (C) 塩基対.

(c) 結晶中での二重らせんDNAのパッキング. 一本の連続したらせんを形成するように重なっている. PHO4と17塩基対DNAオリゴマーの複合体結晶を示した.

現を活性化するとともに, 阻害する場合もある.

真核細胞由来のHTHモチーフをもつ転写因子であるホメオドメイン蛋白質は単量体で回文状ではない塩基配列を認識する. この蛋白質は細胞の発生や分化に関与する重要な転写因子群である. 合計3本の α -らせんで, ホメオドメインと呼ばれるDNA結合ドメイン構造を形成している. 認識配列は数塩基対のAとTの多い配列であるが, 変化に富んでいる. また, 蛋白質と直接に相互作用するDNA部分は, DNAの糖やリン酸基との相互作用を含めても10塩基対以内と短い. このことは, ホメオドメ

イン蛋白質は他の転写因子と協同的にDNAに結合して機能することや, 複数のDNA結合モチーフからなるDNA結合ドメインを同一蛋白質中にもつことと関係していると考えられている. 例えば, 出芽酵母の細胞型を支配する2つのホメオドメイン転写因子MATa1とMAT α 2は, ヘテロ二量体³⁾を形成して半数体に特異的な遺伝子上流に結合して不活化し, 基本数の2倍の染色体をもつ二倍体を保持する (図5(a)). また, ショウジョウバエの体節の分節化に関与するpairedと呼ばれる転写因子のDNA結合ドメインは18残基のループで繋がった2つの

ホメオドメインからなり、3塩基対離れた2つの共通配列 TAATと ATTA を認識する (図5 (b)).⁴⁾ ヒストン H2B 遺伝子の上流で最初に発見された特徴的な8塩基対 (ATTTGCAT) は種々の遺伝子で見出される重要な塩基配列である。この配列を認識する下垂体前葉に特異的な転写活性化因子 Pit-1 や H2B 遺伝子の Oct-1 あるいは線虫の神経細胞分化に関係する転写因子 unc-86 などは POU と呼ばれる DNA 結合ドメインをもつが、これは HTH モチーフをもつ POU 特異的ドメインとホメオドメインとからなる (図5 (c)).⁵⁾ さらに、Oct-1 は単純ヘルペスウイルス由来の因子 VP16 と結合して転写の活性化をすることが知られている。最近報告された酵母由来の RAP1 はクロマチンの末端にある GC 塩基対に富んだテロメアと

表1 DNA との複合体の構造 (DNA/transcription factor complex)

認識部位	モチーフ	例
α-らせん	HTH	phage 434R, phage 434Cro, phage λR, phage λCro, TrpR, LacR, PurR, RAP1, c-Myb*
		homeodomain; engrailed, Antennapedia, paired (homodimer), even skipped, MATα2, MATα1/MATα2 (heterodimer)
		POU, paired
		POU, paired
		POU, paired
		POU, paired
		POU, paired
		POU, paired
		POU, paired
		POU, paired
β-シート	wHTH	CAP, HNF-3γ, PU.1-Ets
	bZip	GCN4, c-Fos/c-Jun
	bHLH	E47, MyoD, PHO4
	bHLHZip	MAX, USF
	Zn-finger	Cys2-His2 type; Zif268, GLI, TTK, GAGA* Zn2-Cys6 type; GAL4, PPRI Cys4 type; GR, ER, RXR/TR, GATA-1*
	β-sandwich	E2, EBNA1
	HMG	LEF-1*, SRY*
	その他	SRF (setum response factor)
	RHH	MetR, ArcR
	その他	TBP, TBP/TFIIA, TBP/TFIIB
ループ	β-sandwich	NF-κB (nuclear factor κB), p53 (tumor suppressor p53)

*はNMRで決定された

図4 バクテリオファージ 434 リプレッサー²⁾ (Bacteriophage 434 repressor)

(a) DNA に結合した HTH モチーフ。α-らせん α2 と α3 が HTH モチーフである。α3 が二重らせん DNA の主溝に沿って入り込み、塩基と相互作用している。
(b) DNA に結合した二量体 434 リプレッサー。二重らせん DNA の同じ側面から主溝に沿って入り込み、塩基と相互作用している。黒く塗ったのが α3 である。

図5 ホメオドメイン (Homeodomain transcription factors)

- (a) DNA に結合した酵母の転写因子 MATα1/MATα2 のホメオドメインのヘテロ二量体。³⁾ MATα2 の C-末端側の α-らせんが MATα1 との会合を安定化している。
(b) DNA に結合したショウジョウバエの転写因子 paired の DNA 結合ドメイン。⁴⁾ 2つのホメオドメインからなる。アミノ末端側に短い2本鎖の反平行 β シートが存在しており、副溝から塩基を認識している。また、ドメイン間のループも副溝で相互作用している。
(c) DNA に結合したヒト転写因子 Oct-1 の POU ドメイン。⁵⁾ POU 特異的ドメインと POU ホメオドメインからなる。ドメイン間のループも副溝に沿って相互作用している。

呼ばれる配列やリボゾーム遺伝子上流に特異的に結合するが、この DNA 結合ドメインもホメオドメインによく似た 2 つの HTH モチーフからなる。⁶⁾ その他、がん原遺伝子産物である転写因子 c-Myb⁷⁾ なども HTH モチーフをもつ。

HTH モチーフの類似のものとして wHTH (winged helix-turn-helix) モチーフがある。これは、二本の α -らせんの間に余分の β -鎖と、HTH モチーフのカルボキシル末端側に β -シートをもつモチーフで、原核細胞と真核細胞の両方に存在する。ヒストン H5 のコアドメインや細胞増殖、分化、発生、癌化などに幅広く関与すると考えられている Ets⁸⁾ などがある。原核細胞では細菌の cAMP 受容体蛋白質である CAP⁹⁾ が有名である。

5. bZip, bHLH, bHLHZip モチーフはヘテロ二量体も形成する

真核細胞に特有で、細胞増殖や癌化などに関与する一連の転写因子に見出された bZip (basic-region leucine-zipper) モチーフ、あるいは筋細胞の分化などに関与する転写因子で発見された bHLH (basic-region helix-loop-helix) モチーフは、塩基配列の認識に直接関与する塩基性領域に蛋白質どうしの会合に関与する疎水性領域が連なっている。これらの転写因子は疎水性領域を通して、同種の分子どうしが会合したホモ二量体や、同じモチーフをもつ異種の分子と会合したヘテロ二量体を形成する能力をもっている。bZip モチーフは、塩基性領域 (b) と会合モチーフであるロイシンが 7 残基おきに現れるロイシン・ジッパー領域 (Zip) が繋がった一本の α -らせんを形成する (図 6 (a))。この Zip 領域ではらせんの二回転ごとに同一の側面に疎水的なロイシンの側鎖が突き出しており、これらの側鎖どうしが会合することにより二量体を形成する (図 6 (b))。認識される基本配列は TGACTCA、または TGACGTCA であり、すべての bZip モチーフに共通している。

bHLH モチーフではロイシン・ジッパー領域の代わりに、二本の α -らせんがループで繋がっており (HLH)、合計 4 本の α -らせんがアミノ末端を同じ方向へ向けて束のように会合する。このループは数残基から数十残基までの長さをもつが、各転写因子で異なっており、転写因子間でこのループを交換すると活性を失う。最近、長いループをもつ出芽酵母のリン酸代謝系の HLH モチーフをもつ転写因子 PHO4 がこのループ部分に短い α -らせんをもつことがわかった (図 6 (c))。変わり種として、bHLH モチーフのカルボキシル末端に Zip 領域をもつ bHLHZip (basic-region helix-loop-helix leucine-zipper) モチーフもある。例えば、癌原遺伝子産物の転写因子である Myc や Max¹¹⁾ などである (図 6 (d))。bHLH モチーフと bHLHZip モチーフに共通した認識基本配列は CAXXTG

図 6 真核細胞に特有な bZip, bHLH および bHLHZip モチーフ (bZip, bHLH and bHLHZip motifs of eukaryotic transcription factors)

- (a) DNA に結合した bZip モチーフをもつ酵母転写因子 GCN4 の DNA 結合ドメイン。¹⁰⁾
- (b) bZip モチーフのロイシン・ジッパーの疎水的相互作用による会合。
- (c) DNA に結合した bHLH モチーフをもつ酵母のリン酸代謝系の転写因子 PHO4 の DNA 結合ドメイン。¹⁾
- (d) DNA に結合した bHLHZip モチーフをもつ転写因子 Max¹¹⁾ の DNA 結合ドメイン。

であるが、種々の変化に富んだ配列が知られている。実際に、bHLHモチーフとbHLHZipモチーフとでは塩基性領域のアミノ酸配列の相同性に差異があり、標的塩基配列とそれを認識するアミノ酸残基にも違いがあることがわかっている。

6. 真核細胞に特有な Zn-finger モチーフは変化に富んだスーパーファミリーを形成している

Zn-finger モチーフとは、亜鉛イオンに配位した約30アミノ酸残基がユニットを形成し、その中の α -らせんがDNAの塩基配列を認識する一群のDNA結合モチーフの総称であり、発見当時、このユニットが複数個連なり、まるで指（フィンガー）のようであるとされたことからこの名前がついた。真核細胞に特有であり、大きく分けて3つの類型がある。ひとつは、2残基ずつのシステインとヒスチジンが配位している型で、フィンガーの繰り返しは多いもので37個にも及ぶ。基本構造は1本の α -らせ

んと2本の β -鎖からなる β -シートとで形成され、亜鉛に配位している2残基のヒスチジンは α -らせんに位置している（図7(a)）。単量体でDNAに結合し、認識配列はGCに富んでいる。一方、Zn-クラスターと呼ばれるDNA結合モチーフは、2つの亜鉛イオンに6残基のシステインが配位したもので、2残基のシステインは2つの亜鉛イオンを架橋している。アミノ末端側の α -らせんが塩基配列を認識する（図7(b)）。グルココルチコイドなどのホルモンの核内レセプターは転写因子であることがわかっているが、これらのDNA結合ドメインは2つの亜鉛イオンを含むZn-fingerモチーフをもつ（図7(c)）。4残基のシステインがそれぞれの亜鉛イオンに配位しており、2つのZn-fingerが折りたたまってひとつのドメインを形成している。アミノ酸配列と三次元構造の相同性が高く、多くの場合、ホモあるいはヘテロ二量体を形成してDNAに結合する。認識配列は3～5塩基対離れたAGAACAもしくはAGGTCA配列で、回文状のものと反復配列のものが

図7 Zn-finger モチーフ (Zn-finger motif)

- (a) マウス由来の転写因子 Zif268¹²⁾ の Cys2-His2 タイプ Zn-finger モチーフ。
- (b) 酵母由来の転写因子 GAL4¹³⁾ の Zn2-Cys6 タイプ Zn-finger モチーフ。
- (c) グルココルチコイドレセプター GR¹⁴⁾ の Cys4 タイプ Zn-finger モチーフ。
- (d) 二量体を形成したグルココルチコイドレセプター GR の DNA 結合ドメインとの複合体。

ある。アミノ末端側の Zn-finger モチーフの α -らせんが塩基配列を認識する (図 7 (d))。転写因子 GATA-1¹⁵⁾ はこのアミノ末端側の Zn-finger モチーフに類似するとされている。

7. 塩基配列は β -シートやループでも認識される

原核細胞には、 β -シートを主溝に挿入して塩基配列を認識する転写因子が知られている。例えば、メチオニンの生合成に関わる遺伝子群を調節するリプレッサーである MetR やサルモネラ菌のバクテリオファージ P22 の ArcR¹⁶⁾ である。これらは β -鎖に 2 本の α -らせんが連なった RHH (ribbon helix-helix) モチーフを持ち、ホモ二量体を形成することにより、この 2 本の β -鎖が反平行 β -シートを形成して、これが塩基配列を認識する。二量体は α -らせんで会合して、さらに二量化する。一方、免疫系やサイトカイン関連遺伝子の調節をする NF- κ B (nuclear factor κ B)¹⁷⁾ や癌抑制遺伝子産物である p53 (tumor suppressor p53)¹⁸⁾ は、ループ領域を主体とした塩基配列の認識をすることが明らかとなった。NF- κ B は、 β -サンドウィッチと呼ばれる β -シートに富んだ 2 つのドメインからなり、両ドメインの側面から突き出したループと両ドメインをつないでいるループが DNA の主溝に巻き付くように結合する。p53 も β -サンドウィッチ構造をもつ。

8. DNA 塩基は配列として読みとられる

転写因子の DNA 結合ドメインと DNA との複合体研究から DNA- 蛋白質相互作用には幾つかの共通点を指摘することができる。ここでは bHLH モチーフをもつ PHO4 と DNA との複合体を例にとって説明する。まず、認識されるコア配列は 6 塩基対 (CACGTG) とされていたが、相互作用の観測された DNA の領域は 12 塩基対に及ぶ (図 8 (a))。特に、リン酸基との水素結合やイオン対は合計 15 個みられ、広い領域に分布している。これらは塩基配列に非特異的な相互作用であるが、DNA 結合への寄与が大きいことを示唆している。コア配列 (6 塩基対であるから合計 12 塩基ある) へは合計 8 本の水素結合がみられた。ここで重要なことは、蛋白質は DNA の塩基を一つ一つ認識しているわけではなく、むしろ配列として読みとられていることである (図 8 (b))。例えば、PHO4 はグアニン塩基 G (1L') のカルボニル酸素 O6 と Arg13A との一本の水素結合を形成することにより、その位置にグアニン塩基が存在する配列と存在しない配列とを識別している (図 8 (b) の C)。このように、グアニン塩基は塩基性アミノ酸であるアルギニンと水素結合を形成することが多い。他の塩基対においても同様で、塩基対当たりに費やされる水素結合は 1 本か 2 本である。また、HLH モチーフの認識配列の中で PHO4 に特有とされていたコア

配列以外の配列、G (4R') と G (5R')、への水素結合あるいはイオン対も 3 本ほど観測された。これらの特異的および非特異的相互作用の数の比は、通常、塩基配列非特異的結合は解離定数が $10^{-4}\text{M} \sim 10^{-6}\text{M}$ であり、特異的結合が $10^{-7}\text{M} \sim 10^{-9}\text{M}$ であるという事実と対応している。ちなみに、常温で、解離定数が一桁変化すると約 6kJ/mol の自由エネルギー変化を伴う。疎水的相互作用も塩基認識に使われる。アラニンのメチル基あるいはセリンやアルギニンのメチレン基とチミンのメチル基のファンデルワールス接触が頻繁に観測される。bZip モチーフをもつ転写因子 Fos と Jun のヘテロ二量体¹⁹⁾ の例を図 8 (c) に示した。

アミノ酸側鎖が塩基の官能基と直接に水素結合を形成する場合が多いが、トリプトファン代謝系遺伝子群のリプレッサーである TrpR とオペレータ DNA やホメオドメイン・DNA 複合体のように高い精度で決定された構造では水分子を介した水素結合による塩基認識が見出されている (後述)。このことは、塩基配列の普遍的な読みとり規則の発見を非常に困難なものにしている。現状では、非常に類縁性の高い DNA 結合ドメインにおいては整合性のある説明が可能であるが、さまざまな構造モチーフに共通する基本メカニズムの概念は非常に貧しい。さらに、構造データを積み上げて、詳細な解析をする必要があるが、非常に複雑ではあるが、筆者には新しいアイデア一つで解決するパズルのようにも思っている。

9. 塩基配列は副溝からも認識される：二重らせん DNA の溝に巻き付く塩基性領域

バクテリオファージ由来の入りプレッサー²⁰⁾ などでは HTH 結合モチーフのアミノ末端側に塩基性アミノ酸残基をもつ部分が存在し、これが DNA 二重らせんの主溝に巻き付いて、塩基を認識するとともにリン酸基と相互作用することが知られていた。また、ホメオドメインにはそのアミノ末端の塩基性ループが副溝から塩基と相互作用するものが見つかっている。例えば、englailed ホメオドメイン²¹⁾ ではアミノ末端の 2 つのアルギニンが副溝から 2 つのチミンと水素結合を形成していた。また、paired ドメインのアミノ末端側には短い 2 本鎖の反平行 β -シートが存在しており、アスパラギン側鎖がグアニンを、グリシンが次のグアニンを、そしてアルギニン側鎖が水分子を介して次のチミンを水素結合によって認識している (図 5 (b))。また、アミノ末端側とカルボキシル末端側の HTH モチーフを結ぶリンカー部分は副溝に沿っており、主鎖を通してチミンやグアニンを認識している。Oct-1 の POU ドメインのリンカー部分も、副溝側でアルギニン側鎖がアデニンと水素結合を形成している。また、Cys2His2 型の Zn-フィンガーである GAGA-12²²⁾ のアミノ末端の塩基性領域が DNA を抱え込むように巻き付き、副

図8 塩基配列の読みとり (Reading out of the DNA sequence)

(a) bHLH モチーフをもつ酵母のリン酸代謝系の転写因子 PHO4 の DNA 結合ドメインと DNA との相互作用。

(b) PHO4 の DNA 結合ドメインによる塩基の認識 (A-H)。

(c) bZip モチーフをもつヒト由来の転写因子 AP-1 のヘテロ二量体 DNA 結合ドメイン¹⁹⁾ による塩基の認識。回文状のコア配列 TGACTCA の半分を図示している。

溝側からアルギニンやリジン側鎖が塩基を認識していることが最近報告されている。GAGA-1 は DNA に巻き付くことにより、DNA をヌクレオソームから剥がす働きを担っていると考えられている。

10. 塩基配列は副溝からも認識される：疎水的相互作用

転写開始部位の上流にある TATA 配列を認識し、転写開始位置を決定する基本転写因子 TBP²³⁾ (TATA ボックス結合蛋白質) は塩基配列特異的に副溝に結合する (図9 (a))。ここでの相互作用は反平行のβシート上の疎水性

アミノ酸残基と塩基や糖とのファンデルワールス接触を主体とした疎水的相互作用であり、水素結合は観測されていない (図9 (b))。また、性決定に関わる SRY や T 細胞受容体の遺伝子の制御に関わる LEF-1²⁴⁾ も、疎水的相互作用により塩基配列特異的に副溝に結合する。これらの転写因子は3本のα-らせんとカルボキシル末端のループからなるL-字形の構造をもち、染色体蛋白質である非ヒストン蛋白質に分類されるHMG (high mobility group) 蛋白質と同じ類型に属するためにHMGドメインと呼ばれ、TTT配列を含むAT塩基対に富む配列を認識する。この場合、アミノ末端側の第1α-らせんが副溝に沿って入

図9 副溝への結合とのDNA湾曲 (The binding to minor groove causes DNA bending or kinking)

(a) DNAに結合した基本転写因子TBP. 複合体結晶構造のDNAオリゴマーへ二重らせんDNAを余分に付加して表してある.

(b) TBPとDNAとの疎水的相互作用. 塩基対の疎水の側面にValやLeuが会合している.

(c) DNAに結合したラクトースオペロンのリプレッサーLacRのDNA結合ドメイン. DNA塩基を認識するHTHモチーフのC-末側にある α -らせんが副溝に挿入されて, DNAの二重らせんを曲げている.

り込み, カルボキシル末端のループは主溝側へ伸びている. 少数ではあるが, 水素結合による塩基認識も含む. 例えば, LEF-1ではアスパラギンとグルタミンの側鎖がグアニンを認識している.

11. 二重らせんDNAの変形と湾曲

転写因子の結合により, DNA構造は変化する. 例えば, PHO4とDNAとの複合体では, 標準的なB型DNAからのずれ (root-mean-square 誤差) はDNAの末端部分で最大6Åもあった (図3(a)). また, 塩基対と糖リン酸骨格部分ではそれぞれ1.5Åと2.0Åのずれがあった. 実際, 塩基配列を認識されている主溝は狭まり, 副溝は反対に広がっている. さらに大きな構造変化がいくつも観測されている. 特に, 前項で紹介した疎水的相互作用による副溝へ結合は大きなDNAの構造変化を引き起こす. TBPは110°の湾曲と18Åのらせん軸のずれを引き起こす (図9(a)). また, SRYやLEF-1は疎水的な α -らせんを副溝に密着させてDNAを70°~120°も折り曲げることが知られている. これらの認識される塩基配列はAT塩基対に富む点が共通している. DNAを折り曲げることにより, 他の転写因子の近隣の配列への結合や転写因子間の相互作用を促進する機能を果たしていると考えられている. さらに, 原核細胞由来の転写因子としては, 大腸菌のプリン代謝あるいはラクトース代謝系のリプレッサーPurR²⁵⁾やLacR²⁶⁾はHTHモチーフで塩基配列を認識するが, さらに, このモチーフのカルボキシル側にある余分の α -らせんが副溝に挿入されてDNAを約45°折り曲げる (図9(c)). このように, 副溝に結合してその溝を拡げることが, 大きならせんの曲がりを生ずるひとつのメカニズムとして指摘できる. その他, 主溝に結合してDNAを大きく折り曲げる例として, 前出の大腸菌のCAP (wHTHモチーフ)や酵母のMATa1/MAT α 2 (HTHモチーフ), あるいはパピローマウイルスのE2²⁷⁾やEpstein-BarrウイルスのEBNA1²⁸⁾などがある. 後者の2つは二量体として β -サンドウィッチを形成して露出した α -らせんで塩基配列を認識するが, 複製の開始原点に結合するという共通の機能をもつ点興味深い. DNA二重らせんの変形メカニズムは少数の構造パラメータを用いた構造力学的解析が可能である. この観点からの蛋白質結合によるDNA二重らせんの変形は他稿で解説しているので, 参考されたい.²⁹⁾

12. 水分子を介した水素結合による塩基認識もある

蛋白質もDNAも分子表面には水和水を保持している. 溶液中での表面水和水の滞留時間は500ps程度であり, バルクの水分子との交換は早い. しかしながら, 分子表面のくぼみに入り込んだ水和水の滞留時間は15nsから200 μ sと比較的長い. これらの中には蛋白質やDNAの一

図10 大腸菌トリプトファンリプレッサー TrpR でみられた水分子を介した塩基認識 (Water-mediated base recognition observed in TrpR/DNA complex)

部としてふるまう水和水もあるにちがいない。転写因子とDNAとの複合体でも、相互作用界面に水分子が見つかった。例えば、434 リプレッサー、TrpR、HNF-3 γ 、MATa1/MAT α 2, paired, RAP1, E2 などである (図10)。また、制限酵素 BamHI³⁰⁾ でも見つかった。TrpR とDNAとの複合体ではすべての塩基との水素結合が水分子を介した水素結合であり、直接の水素結合は観測されなかったことは特筆すべきである。また、MATa1/MAT α 2 とDNAとの複合体では主溝と副溝で水分子を介した多くの水素結合が見出された。これらの水分子は、標的DNA単体の結晶でのDNAの水和水や、TrpR単体の結晶での蛋白質の水和水の位置と比較検討する必要があるが、TrpRの場合、単体DNAの主溝に存在する水和水の大部分が複合体でも存在し、水分子を介した水素結合もこれらの位置の水分子を利用していることがわかっている。³¹⁾

13. 非対称結合と共通配列以外の塩基対認識

二量体を形成してDNAに結合する転写因子の場合、原核細胞由来では塩基配列も広い範囲にわたって二回対称をもち、結合はほぼ対称的となる。一方、真核細胞由来の転写因子の場合、ホモ二量体を形成してDNAに結合する場合であっても、一方に結合が偏ることがある。例えば、PHO4の認識配列は対称性のある共通配列 (E-box) としては6塩基対であるが、相互作用領域は20塩基対以上にもおよび、共通配列から3'-側 (下流) へ大きくずれる。これはDNA結合ドメインのみを用いても同様の傾向を示す。実際に、共通配列以外の塩基対で結合に重要であるものが判っており、それらは3'-側 (下流) に多い。この非対称結合の起源は何であろうか？ DNA結合ドメインのDNAとの接触面とDNA表面の形状が必ずしも対称的結合が最も適合するようには設計できておらず、DNA結合ドメインは少しでも広い接触面を確保するためにずれようとするが、そこで安定な相互作用の形成され

る塩基配列側へずれると考えることができる。

ヘテロ二量体が対称性をもった共通配列を認識するときはどうだろうか？例えば、bZip転写因子であるc-Jun/c-Fosヘテロ二量体 (これをAP-1と呼ぶ) の場合、結晶構造中ではc-JunがDNAの下流側になる結合とc-FosがDNAの下流側になる結合の両方が得られている。多分、溶液中では両方の結合が平衡関係にあると推察される。それでは、どちら側が遺伝子発現活性化の機能を発揮するのであろうか？Fe (II)-EDTA錯体をc-Junあるいはc-Fosのアミノ末端に付加してDNAに結合させると、この錯体は近傍のDNA鎖を切断する。この切断部位からc-Junまたはc-Fosがどちらに結合したかがわかる。AP-1は他の転写因子、例えばもともとT細胞を活性化する核因子である転写因子NFATと協同的にDNAに結合し、機能する。そこでNFATとAP-1の結合部位を含むDNAを用いて実験してみた結果、c-Junが対称配列のNFAT側の半分を、c-Fosがその逆側を認識していることが明らかになった。³²⁾

14. 制御ドメインと転写活性化ドメイン

転写因子としての活性は、補助因子の結合、リン酸化、分解あるいは上述のように転写因子どうしのヘテロ二量体形成を通して制御される。最も一般的な補助因子はホルモン、ビタミン、cAMPなどの低分子化合物であるが、これらの結合は転写因子としてよりも広く受容体として理解すべきであり、そこには構造変化の誘起による機能変換というアロステリックと呼ばれる一般的な現象がある。これについては、原核細胞由来で比較的小さい転写因子であるCAP、TrpRあるいはLacRで議論されている (図11)。活性化を通した転写制御が一義的な真核細胞においては、転写因子として最も特徴的な転写活性化ドメインの構造の解明は大きな一里塚となるであろうが、残念ながら解析例がない。アミノ酸配列の特徴としては酸性残基、プロリン残基、あるいはセリンやスレオニン残基のクラスターの存在などが指摘されている。

15. 転写因子間の接触部位

遺伝子発現制御機構を考えると転写因子間の相互作用あるいは協同的結合や機能発現は極めて重要であり、特に制御ネットワークの複雑な真核細胞においては本質的な問題である。上記のNFATとAP-1では、接触部位はNFATのDNA結合ドメインのループ領域とc-Junの同じくDNA結合ドメインの塩基性領域とそれに近いZip領域であることが示唆されている。³³⁾ また、bZip様のDNA結合ドメインをもつ転写因子MafBはEts-1に結合してその転写活性化能を阻害するが、両者の結合部位もDNA結合ドメインにあることが示されている。³⁴⁾ さらに、wHTHをもつOmpRの転写活性化は、基本転写因子でありRNA

(a)

(b)

図 11 全長蛋白質としての転写因子の構造 (A whole structure of transcription factor)

(a) DNA に結合したラクトースオペロンのリプレッサー LacR. 二量体 LacR がさらに二量体を形成して 2 本の DNA オリゴマーに結合している。

(b) LacR のドメイン構成の模式図と制御メカニズム。ラクトース (▲) が結合した二量体 LacR は N-末端側のドメイン間にイオン対を形成して DNA 結合ドメイン (楕円) どちらの会合が阻害されて DNA 結合の能を失う (左図)。

ポリメラーゼのサブユニットの一つである α -サブユニットに働くが、実際、両者の接触部位も DNA 結合ドメインにあることがわかった。³⁵⁾ このように、転写因子や基本転写因子間の相互作用は、従来、制御ドメインや転写活性化ドメインを通して起こると考えられてきたが、最近の研究は、むしろ DNA 結合ドメインを通して起こっていることを示唆している。複数の転写因子や基本転写因子が協同的に DNA に結合した複合体の結晶解析は、真核細胞の基本転写因子である TFIIB や TFIIA を含んだ TBP/TFIIB³⁶⁾ あるいは TBP/TFIIA³⁷⁾ と DNA との複合体などがあるが、まだ始まったばかりである。紙面の都合で割愛するが、これらの巨大な複合体は molecular assem-

bly と呼ばれており、現代構造生物学の象徴的な標的である。

16. 展望

創生期の分子生物学の精神と構造生物学の精神とは本質的に同一である。構造分子生物学 (外国では molecular and structural biology などと記す) とでも呼ぶべきこの方向の研究は、今や、細胞生物学の荒野を開拓しつつある。新しいパラダイムが次々と発見されており、すでに 15 年以上も *Nature*, *Cell*, *Science* といった生物科学の top journals を席卷し続けている。これはブームなどという生半可なものではなく、まさに生物学の一大変革である。

文 献

- 1) T. Shimizu, A. Tsumoto, K. Ihara, M. Shimizu, Y. Kyogoku, N. Ogawa, Y. Oshima and T. Hakoshima: *EMBO J.* **16**, in press (1997) .
- 2) A. K. Aggarwal, D. W. Rodgers, M. Drott, M. Ptashne, and S. C. Harrison: *Science* **242**, 899-907 (1988) .
- 3) T. Li, L. T. Stark, M. R. Johnson, and A. D. Wolberger: *Science* **270**, 262-269 (1995) .
- 4) W. Xu, M. A. Rould, S. Jun, C. Desplan, and C. O. Pabo: *Cell* **80**, 639-650 (1995) .
- 5) J. D. Klemm, M. A. Rould, R. Aurora, W. Herr and C. O. Pabo: *Cell* **77**, 21-32 (1994) .
- 6) P. Konig, R. Giraldo, L. Chapman and D. Rhodes: *Cell* **85**, 125-136 (1996) .
- 7) K. Ogata, S. Morikawa, H. Nakamura, A. Sekikawa, T. Inoue, H. Kanai, A. Sarai, S. Ishii and Y. Nishimura: *Cell* **79**, 639-648 (1994) .
- 8) R. Kodandapani, F. Pio, C. -Z. Ni, G. Piccialli, M. Klemsz, S. McKercher, R. A. Maki and K. R. Ely: *Nature* **380**, 456-460 (1996) .
- 9) S. C. Schultz, G. C. Shields and T. A. Steitz: *Science* **254**, 1001-1007 (1991) .
- 10) T. E. Ellenberger, C. J. Brandl, K. Struhl and S. C. Harrison: *Cell* **71**, 1223-1237 (1992) .
- 11) A. R. Ferre-D'Amare, G. C. Prendergast, E. B. Ziff and S. K. Burley: *Nature* **363**, 38-45 (1993) .
- 12) N. P. Pavletich and C. O. Pabo: *Science* **252**, 809-817 (1991) .
- 13) R. Marmorstein, M. Carey, M. Ptashne and S. C. Harrison: *Nature* **356**, 408-414 (1992) .
- 14) B. F. Luisi, W. X. Xu, Z. Otsinowski, L. P. Freedman, K. R. Yamamoto and P. B. Sigler: *Nature* **352**, 497-505 (1991) .
- 15) J. G. Omichinski, G. M. Clore, O. Schaad, G. Felsenfeld, C. Trainor, E. Appella, S. J. Stahl and A. M. Gronenborn: *Science* **261**, 438-446 (1993) .
- 16) B. E. Raumann, M. A. Rould, C. O. Pabo and R. T. Sauer: *Nature* **367**, 754-757 (1994) .
- 17) C. W. Muller, F. A. Rey, M. Sodeoka, G. L. Verdine and S. C. Harrison: *Nature* **373**, 311-317 (1995) .
- 18) Y. Cho, S. Gorina, P. D. Jeffrey and N. P. Pavletich: *Science* **265**, 346-355 (1994) .
- 19) J. N. Glover and S. C. Harrison: *Nature* **373**, 257-261 (1995) .
- 20) N. D. Clarke, L. J. Beamer, H. R. Goldberg, C. Berkower and

- C. O. Pabo: *Science* **254**, 267-270 (1991) .
- 21) C. R. Kissinger, B. Liu, E. Martin-Blanco, T. B. Kornberg, and C. O. Pabo: *Cell* **63**, 579-590 (1990) .
- 22) J. G. Omichinski, P. V. Pedone, G. Felsenfeld, A. M. Gronenborn and G. M. Clore: *Nature Struct. Biol.* **4**, 4122-132 (1997) .
- 23) J. L. Kim, D. B. Nikolov and S. K. Burley: *Nature* **365**, 520-527 (1993) ; Y. Kim, J. H. Geiger, S. Hahn and P. Sigler: *Nature* **365**, 512-520 (1993) .
- 24) J. J. Love, X. Li, D. A. Case, K. Giese, R. Grosschedl and P. E. Wright: *Nature* **376**, 791-795 (1995) .
- 25) M. A. Schumacher, K. Y. Choi, H. Zalkin and R. G. Brennan: *Science* **266**, 763-770 (1994) .
- 26) M. Lewis, G. Chang, N. C. Horton, M. A. Kercher, H. C. Pace, M. A. Schulacher, R. G. Brennan and P. Lu: *Science* **271**, 1247-1254 (1996) .
- 27) R. S. Hegde, S. R. Grossman, L. A. Laimins and P. B. Sigler: *Nature* **359**, 505-512 (1992) .
- 28) A. Bochkarev, J. A. Barwell, R. A. Pfuetzner, E. Bochkareva, L. Frappier and A. M. Edwards: *Cell* **84**, 791-800 (1996) .
- 29) 箱嶋敏雄, シリーズ・ニューバイオフィジックス3「構造生物学とその解析法」, 印刷中, (日本生物物理学会編) 共立出版社 (1997年) .
- 30) M. Newman, T. Strzelecha, L. Dorner, I. Schildfraut and A. K. Aggarwal: *Science* **269**, 656-663 (1995) .
- 31) Z. Shakked, G. Guzikovich-Guerstein, F. Frolow, D. Rabinovich, A. Joachimiak and P. B. Sigler: *Nature* **368**, 469-473 (1994) .
- 32) L. Chen, M. G. Oakley, J. N. M. Glover, J. Jain, P. B. Dervan, P. G. Hogan, A. Rao and G. L. Verdine: *Curr. Biology* **5**, 882-889 (1995) .
- 33) S. A. Wolfe, P. Zhou, V. Dotsch, L. Chen, A. You, S. N. Ho, G. R. Crabtree, G. Wagner and G. L. Verdine: *Nature* **385**, 172-176 (1997) .
- 34) M. Sieweke, H. Tekotte, J. Frampton and T. Graf: *Cell* **85**, 49-60 (1996) .
- 35) H. Kondo, A. Nakagawa, J. Nishihira, Y. Nishimura, T. Mizuno and I. Tanaka: *Nature Struct. Biol.* **4**, 28-31 (1997) .
- 36) D. B. Nikolov, H. Chen, E. D. Halay, A. A. Usheva, K. Hisatake, D. K. Lee, R. G. Roeder and S. K. Burley: *Nature* **377**, 119-128 (1995) .
- 37) S. Tan, Y. Hunziker, D. F. Sargent and T. J. Richmond: *Nature* **381**, 127-134 (1996) ; J. H. Geiger, S. Hahn, S. Lee and P. Sigler: *Science* **272**, 830-840 (1996) .

プロフィール

箱嶋敏雄 Toshio HAKOSHIMA

奈良先端科学大学院大学 バイオサイエンス研究科, 〒630-01 生駒市高山町8916-5

Department of Molecular Biology, Nara Institute of Science and Technology

TEL. 0743-72-5570 FAX. 0743-72-5579

E-mail: hakosima@bs.aist-nara.ac.jp

最終学歴：大阪大学大学院 薬学研究科 博士課程
専門分野：構造生物学, 構造薬理学, 分子生物物理学

現在のテーマ：分子複合体の構造解析を通じた転写制御系および細胞内情報伝達系での分子認識と医学への応用