コラーゲンの比較生化学

吉 里 勝 利*

1. はじめに

多細胞動物は、多数個の細胞が集合して、一個の個体として生活している。分類学的には、海綿動物門を最も進化の程度が低い体制の多細胞動物とし、脊椎動物門を頂点とする系統樹が考えられている。両者に挟まれて多種多様の動物が存在する。これら多細胞動物の起源については、動物性原生生物(原生動物: protozoa)と考えるのが自然であろうが科学的証拠はない。

多細胞動物は、即ち後生動物(metazoa)であるが、これから中生動物(mesozoa)と側生動物(parazoa)を除いたものを真正後生動物(eumetazoa)とし、これに腔腸動物から脊椎動物に至る全ての動物を含める分類法が一般的である。つまり腔腸動物以上の真正後生動物は、“真正多細胞動物”として一括して考えてよい。

後生動物の特徴の一つは、構成細胞が機能的、形態的に分化しており、同種の細胞が有機的に集合して、種々の組織を形成することから、“組織動物”とも呼ばれる。

後生動物が原生動物のある種を祖先形として生じたであろうことは既述したように疑いはないが、それでは、どの原生動物が祖先となったのであろうか。一つの有力な考え方は、ボルボックス(*Volvox*)などの鞭毛虫群体に祖先形を求めるものである。群体の形成過程と海綿動物の個体発生との類似性、群体の構成個員の間に栄養個虫と生殖個虫の分化があること、後生動物の精子が鞭毛虫の形態をしていることなどがその論拠となっている。

生命の進化を考える時、二つの画期的出来事があったに違いない。一つは化学進化から細胞が出現したことであり、他は、単細胞生命の多細胞化である。筆者は特に後者に興味をもっている。多細胞動物化の仕組みに直接関係ある資料をもっているわけでは

ないが、この問題に迫る一つのアプローチを本稿で紹介したい。

2. コラーゲンをプローブとする

“組織動物”が細胞を組織化する時に必要な物質として二種類考えられる。一つは細胞同士を結合させる物質(細胞接着分子, cell adhesion molecule)であり、他は細胞の足場となる物質(基質, substrate)である。これに細胞は接着し、集合化が可能になる。後者は特に細胞外基質物質(extracellular matrix substance; ECM)と呼ばれる。単細胞動物の多細胞化の過程が、生物の進化の一つの過程としてあったとするならば、これら二種類の物質をコードしている遺伝子の出現は、この過程にとって重要な意味をもつ。これらの物質をプローブとすることによって単細胞動物から多細胞動物への進化の機構を分子レベルで研究する道が拓かれるだろう。

筆者は、これら二種類の物質の中でも後者 ECM に特に注目している。ECM には、大別すると三種類知られているが、その中でもコラーゲンは、結合組織の主要タンパク質として、多細胞動物の体制を物質的に論ずる時、無視できない重要なものである。

本稿ではこのコラーゲン分子に注目して、多細胞体制の起源の問題を論じる。本稿に関する研究は、世界的にみてもまだ未発達分野であり、系統だてて論ずるための資料が非常に乏しい。そこで本稿では、多細胞動物体制の進化をコラーゲン分子およびその遺伝子の性質から考え、その外挿的情况としての起源の問題を推理してみることにする。

3. コラーゲン分子

タンパク質としてのコラーゲン分子の特徴は、三本鎖からできているということと、その三本鎖がコラーゲンらせん構造と呼ばれる特異的な立体構造を

*Katsutoshi YOSHIKATO, 広島大学理学部動物学教室(〒730 広島市中区東千田町 1-1-89)

図1 コラーゲンのドメイン構造¹⁾ □, トリプルヘリカルドメイン; ▨, 棒線の所で Gly-X-Y リピート構造に中断があることを示す; ■, 球状(ノンヘリカル)ドメイン; ▤, シグナルペプチド; N-P, N-プロペプチド; N-TP, N-テロペプチド; TH, トリプルヘリックス; C-TP, C-テロペプチド; C-P, C-プロペプチド

もつことである。三本鎖の一本一本を α 鎖と呼ぶ。 α 鎖の全てが同一のものはホモトリマーであり $[\alpha_1]_3$ で表わされる。ヘテロトリマーには二種類あり, $[\alpha_1]_2\alpha_2$ か $\alpha_1\alpha_2\alpha_3$ である。脊椎動物コラーゲンは現在までに13種類知られており, それぞれ I 型, II 型, ... XIII 型と呼ぶ(表1)。 α 鎖は, 型によって異なる。脊椎動物の体に最も多量に存在する I 型コラーゲンは $[\alpha_1(I)]_2\alpha_2(I)$ と表記される。

立体構造の特徴で分類すると, 線維性コラーゲンと非線維性コラーゲンに分けることができる(表1)。この分類法はコラーゲン分子の進化を考える時重要である。

コラーゲンのアミノ酸シーケンスの特徴は, Gly-X-Y の三つのアミノ酸の配列が一つの単位となっていて, この単位がくり返されることである。X, Y の位置には, それぞれプロリンとヒドロキシプロリンがくることが多く, その確率はそれぞれ34%と33%である。(Gly-X-Y)_n のシーケンスの部位をコラーゲンドメインと呼び, ここはコラーゲンらせん構造を示す部域(ヘリカルドメイン)である。他の部分は非コラーゲンドメインで, らせん構造を示さない。

線維性コラーゲンの代表として I 型コラーゲン, 非線維コラーゲンの代表として IV 型コラーゲンのドメイン構造を図1に示す。I 型コラーゲンの特徴は, トリプルヘリカルドメインが大部分を占めていることである。1000残基あまりのアミノ酸が Gly-X-Y のくり返しで連続して繋がっている。N末端側にシ

グナルペプチド, N末プロペプチドおよびN-テロペプチドのドメインが, またC末端側にC-テロペプチドおよびC末プロペプチドがある。シグナルペプチドとN末およびC末プロペプチドドメインは, 翻訳後プロセッシングを受けて切断される。

IV型コラーゲンは, 中心部のトリプルヘリカルドメイン, C末端の球状ドメイン(NC1)およびN末端の球状ドメイン(NC2)からなる。線維性のコラーゲンと異なり翻訳後のプロセッシングはない。I型のヘリカルドメインが Gly-X-Y リピートが連続しているのに比べて, IV型では, このくり返しの中断が20回以上もある。

4. コラーゲン遺伝子

線維性コラーゲン(I型 $\alpha 2$ 鎖)の遺伝子構造を図2に示す。52のエキソンから成り, エキソン7~48がトリプルヘリカルドメインをコードしている。エキソン6は, N末プロペプチドの最後の4個のアミノ酸, N末テロペプチドの12個のアミノ酸およびトリプルヘリカルドメインの初めの3個のアミノ酸をコードしているので, joining exon と呼ばれる。同様の joining exon がC末側にもある(エキソン49)。エキソンとイントロンの配置を図3に示す。

トリプルヘリカルドメインは44個のエキソンからなる。これから joining exon 2個を除いた42個(エキソン7~48)のエキソンサイズを表2に示した。42個のうち23個(55%)は54bp(塩基対), 8個(19%)は108bp, 1個(2%)は162bp, 5個(12%)は45bp,

表1 コラーゲン一家¹⁾

型	α 鎖	遺伝子座	主な α 鎖組成	超分子構造
I	$\alpha 1(I)$	COL1A1	$[\alpha 1(I)]_2 \alpha 2(I)$,	線維性
	$\alpha 2(I)$	COL1A2	$[\alpha 1(I)]_3$	
II	$\alpha 1(II)$	COL2A1	$[\alpha 1(II)]_3$	線維性
III	$\alpha 1(III)$	COL3A1	$[\alpha 1(III)]_3$	線維性
IV	$\alpha 1(IV)$	COL4A1	$[\alpha 1(IV)]_2 \alpha 2(IV)$	非線維性
	$\alpha 2(IV)$	COL4A2		
	$\alpha 3(IV)$	COL4A3		
	$\alpha 4(IV)$	COL4A4		
	$\alpha 5(IV)$	COL4A5		
V	$\alpha 1(V)$	COL5A1	$[\alpha 1(V)]_2 \alpha 2(V)$,	線維性
	$\alpha 2(V)$	COL5A2	$\alpha 1(V) \alpha 2(V) \alpha 3(V)$	
	$\alpha 3(V)$	COL5A3		
VI	$\alpha 1(VI)$	COL6A1	$\alpha 1(VI) \alpha 2(VI) \alpha 3(VI)$	非線維性
	$\alpha 2(VI)$	COL6A2		
	$\alpha 3(VI)$	COL6A3		
VII	$\alpha 1(VII)$	COL7A1	$[\alpha 1(VII)]_3$	非線維性
VIII	$\alpha 1(VIII)$	COL8A1	不明	非線維性
IX	$\alpha 1(IX)$	COL9A1	$\alpha 1(IX) \alpha 2(IX) \alpha 3(IX)$	非線維性
	$\alpha 2(IX)$	COL9A2		
	$\alpha 3(IX)$	COL9A3		
X	$\alpha 1(X)$	COL10A1	$[\alpha 1(X)]_3$	不明
XI	$\alpha 1(XI)$	COL11A1	$\alpha 1(XI) \alpha 2(XI) \alpha 3(XI)$	線維性
	$\alpha 2(XI)$	COL11A2		
	$\alpha 3(XI)$	COL2A1		
XII	$\alpha 1(XII)$	COL12A1	$[\alpha 1(XII)]_3$	非線維性
XIII	$\alpha 1(XIII)$	COL13A1	不明	不明

図2 I型 α 2(ニワトリ)の遺伝子²⁾ 1~52はエクソン番号；■，C末又はN末プロペプチドのエクソン；▨，C末又はN末のテロペプチドのエクソン；▨，Gly-X-Y リピー特構造のエクソン。エクソン1の読み始めコドンより3'側はシグナルペプチドをコードしている。

図3 I型の α 2の遺伝子のエクソン-イントロン配置図²⁾ 黒く塗りつぶしてある部分がエクソン，その間の横棒がイントロンを表わす。5'→3'に描いてある。下の目盛りは塩基対(bp)数をkbで表わしている。

また残りの5個(12%)は99bpからなっている。このように全てのトリプルヘリカルエクソンのサイズは、9bpの倍数である：45bpは5個のGly-X-Yトリプレット，54bpは6個，99bpは11個，108bpは12個，162bpは18個のトリプレットのコドンにそれぞれ対応している。

エクソンの大半は54bpであることから，これが単位となって増幅してできたものが先祖型コラーゲン遺伝子であろうと考えられている。108と162bpのエクソンは介在イントロンが欠失することによってできたもの，また45bpと99bpのエクソンは，2個のエクソン間の組み換えでできたものと考えられている。

C末プロペプチドは， α 鎖が会合してヘリックス構造を形成する時に重要な働きをする。ヘリックス形成にさきだって，鎖内および鎖間にS-S結合ができる。このS-S結合が正しいヘリックス形成に

重要である。C末プロペプチド内でのシステイン残基およびその周辺のアミノ酸シーケンスは動物種にかかわらずよく保存されている。従ってシステインの位置と数は，コラーゲンの比較生化学的研究にとって重要な情報となる。

IV型コラーゲン(非線維性)のエクソンのサイズを表3に示す。線維性コラーゲンはかなり異なった分布と示している。しかし，Gly-X-Yリピートの中断やsplit codonによる補正を行うと9-bpが基本になっていることがわかる。

5. 線虫のコラーゲン遺伝子

線形動物門(Nematoda)に属する線虫 *Caenorhabditis elegans* の表皮(cuticle)は下皮(hypodermis)の細胞分泌物で，主要成分はコラーゲンである。コラーゲンの遺伝子は50~100個存在する。コラーゲン分子は分子量が小さいことが特徴で30kD ぐらい

表2 線維性コラーゲンのトリプルヘリカルドメインのエキシンの構造¹⁾

Exon no.	Size(bp)	Exon no.	Size(bp)	Exon no.	Size(bp)
7	45	21	108	35	54
8	54	22	54	36	54
9	54	23	99	37	108
10	54	24	54	38	54
11	54	25	99	39	54
12	54	26	54	40	162
13	45	27	54	41	108
14	54	28	54	42	108
15	45	29	54	43	54
16	54	30	45	44	108
17	99	31	99	45	54
18	45	32	108	46	108
19	99	33	54	47	54
20	54	34	54	48	108

表3 Gly-X-Y リピートに対するエキシンの サイズ分布¹⁾

Gene	27	36	45	54	63	72	81	90	99	108	117	126	135	144	153	162	171	189
$\alpha 2(I)$	0	0	5	23	0	0	0	0	5	8	0	0	0	0	0	1	0	0
$\alpha 1(IV)-1$	2	2	3	2	2	1	1	3	3	0	0	0	0	0	0	0	0	0
-2	0	0	1	2	0	2	0	0	1	1	1	1	1	1	1	0	2	1

gene の数字は塩基対 (bp) 数を表す。

$\alpha 1(IV)-1$ は中断による補正をほどこしていない実数

$\alpha 1(IV)-2$ は補正後の数

である。

遺伝子構造から明らかにされている6個のコラーゲン分子 (Col-1, Col-2, Col-8, Col-19, Col-6, Col-14) のドメイン構造を図4に示す。図にはシステインの位置も示してある。脊椎動物に比べて Gly-X-Y リピート中断箇所が多くて短いことが特徴である。最近IX型の脊椎動物軟骨コラーゲンの遺伝子構造が明らかにされた⁴⁾。それによれば Gly-X-Y のドメインは短く100~300個のアミノ酸に相当する。このヘリカルドメインは、システイン残基を含む18

~84個のアミノ酸を含むノンヘリカルドメインによって囲まれている。線虫のコラーゲンは、脊椎動物のIX型コラーゲンに似ていると考えられる。

線虫の突然変異遺伝子として知られていた dpy-13⁵⁾ と sqt-1⁶⁾ の遺伝子構造があいついで明らかにされたが、これらはいずれもコラーゲン遺伝子である (図5)。これらの構造も基本的には、図4に示したものと同じである。図5には、イントロンの位置とサイズも示したが、dpy-13では2個、sqt-1では1個しかなく、これも脊椎動物のものとは、異なっ

図4 線虫のコラーゲンのドメイン構造³⁾ [斜線], Gly-X-Y リピートドメイン；—, ノンヘリカルドメインC；システインの位置；ATG, 読み始めコドン；TAA (TAG), 読み終りコドン。数字はアミノ酸残基数。

図5 線虫のコラーゲン遺伝子(dpy-13及びsqt-1) [斜線], ノンヘリカルドメイン；[斜線], Gly-X-Y リピートドメイン；—, インtron。数字は塩基対数。

ている。

6. ショウジョウバエのコラーゲン遺伝子

昆虫ではキイロショウジョウバエ *Drosophila melanogaster* のコラーゲン遺伝子の構造が明らかにされている。哺乳類のIV型に属すると考えられる。

両者のα鎖をコードしている遺伝子構造はいくつかの類似的性質を示す。Gly-X-Y リピートの中断部位21箇所のうち11の位置はよく近似している。しかしエクソンとイントロン構造は異なっている。ショウジョウバエは8個のエクソンと7個のイントロンを含むが、ヒトでは、52個のエクソンと51個のイン

図6 カイメンコラーゲン EmC4のドメイン構造 □, ヘリカルドメイン; —, ノンヘリカルドメイン; ▲, システインの位置。数字はアミノ酸の数。

トロンを含んでいる。

7. カイメンのコラーゲン遺伝子

淡水産カイメン *Ephydatia miilleri* のコラーゲン遺伝子の構造が明らかにされている⁷⁾。8個のエキシソンのシーケンスが決定されている。そのうち2個は54bp からなり、他の3個はその倍数(108, 162, および270)の塩基対からなっている。一個は57bp からなるが、これはその一部に Gly-Ser-Ile-Arg の不完全配列をもつためである。ヘリカルドメインのC末端の2つのエキソンは126bp と18bp からなっている。18bp および270bp は、脊椎動物のエキソンにはみられないサイズである。非常に短いサイズのエキソンと非常に長いサイズのエキソンが混在しているのは興味深い。C末端のプロペプチドのエキソンのアミノ酸シーケンスを比較してみると、Ⅱ型のコラーゲンに相溶性が高いことが指摘されている。

同じ淡水産カイメンにこれとは別種のコラーゲン遺伝子の構造が報告されている⁸⁾。芽球から発生した3日目のスポンジから poly(A)-rich RNA を得た後、これを用いて無細胞系でのタンパク合成を行わせた。この実験により、カイメンは分子量200kDa, 160kDa, 81kDa および48kDa のコラーゲンを合成していることがわかった。また得られた cDNA ライブラリーからヒト $\alpha 1(I)$ プロコラーゲン cDNA をプローブとしてこれに相補的なクローンを選したところ、48kDa コラーゲンに対応すると考えられる陽性クローン EmC4 が得られた。この遺伝子は48K のうちN末側の35.7K をコードしている。EmC4 は、図6に示すようにN末側に23個のアミノ酸からなるコラーゲンヘリカルドメイン、16個のアミノ酸(その内3個はシステイン)からなるノンヘリカルドメイン、171個のコラーゲンドメイン、9個のシステインを含む156個のC末プロペプチドドメインの4つのドメイン構造をもつタンパク質をコー

ドしている。

EmC4のヌクレオチドシーケンスの特徴を調べてみると、C末端のノンヘリカルプロペプチドドメインに2箇所Ⅳ型コラーゲンおよび線虫のCol-1, Col-2 dpy-13に近似のシーケンスが存在する。EmC4は、従って、非線維性コラーゲンに属するものと考えられる。

8. ウニのコラーゲン遺伝子

ウニのコラーゲン分子を単離精製して、その α 鎖分析がイイジマクロウニとバフンウニで行われている⁹⁾。いずれも外殻のコラーゲンである。このコラーゲンの抗体は、他種ウニのコラーゲンやヒトデのコラーゲンとは反応するが、哺乳類のⅠ型、Ⅱ型、Ⅲ型、Ⅳ型、Ⅴ型とは全く反応しないかあるいは弱い反応性しか示さない。4本の α 鎖を含み、少なくとも二種のコラーゲン分子が存在すると考えられる。この成体のコラーゲンに対する抗体は、幼生の骨片に反応する。

Paracentrotus lividus のプリズム胚から RNA を抽出し cDNA ライブラリーを作製後、線虫コラーゲン Col-1 に相補的なクローン (COLL 1 α) が選択された¹⁰⁾。Gly-X-Y のリピート部位が1434bp、C-プロペプチドドメインが756bp でそのうち72bp はテロペプチド部位に相当する。

COLL 1 α の遺伝子構造を明らかにするためゲノムライブラリーで解析が行われ、プロペプチド部位のエキソン-イントロンの配置がわかった(図7)。52個のエキソンの構造が明らかになっている。C-プロペプチドは5つのエキソン(48-52)に分散している。このことは、脊椎動物が4つのエキソン(49-52)からなっていることと異なっている。後者の場合49エキソンが junction exon であるが、ウニではこれが48と49エキソンに分離しているものと考えられている。COLL 1 α 遺伝子は、脊椎動物の線維性コラーゲンに相当し、そのヘリカルドメインのエ

図7 ウニのコラーゲン COL1 α の遺伝子構造 , ヘリカルドメインのエキソン; , カルボキシプロペチドのエキソン; ★, システインコドン位置。上はウニの, 下は脊椎動物のコラーゲンを表わしている。遺伝子中の数字はエクソンのサイズ(bp)を, また()付きは平均値を示している。

キソンは54-bp をモチーフとして作られている。

ムラサキウニ (*Strongylocentrotus purpuratus*) では, 脊椎動物のIV型コラーゲンに類似のコラーゲン遺伝子の一部がクローニングされている¹¹⁾。このクローニングにはマウスの $\alpha 1$ (IV) コラーゲンの cDNA がプローブとして用いられた。Gly-X-Y リピードが212bp で一箇所このリピードの中断がある。

これらのことから, ウニ類でも, 線維性コラーゲンと非線維性のコラーゲンの遺伝子が存在していることがわかる。

9. 中生動物と単細胞生物のコラーゲン遺伝子

中生動物として分類される二胚虫類 (Dicyemida) にコラーゲンが存在するかは興味をもたれるところである。マダコに寄生する二胚虫 *Dicyema acuticephalum* を乾重にして 4 mg 集めヒドロキシプロリンを分析したが検出限界以下であった。また 40 μ g の二胚虫タンパク質を SDS-PAGE で分離して得られたバンドの中にはコラーゲナーゼ消化性のものはなかった。

原生動物にコラーゲン遺伝子があるかという観点からの研究報告は筆者の知る限りない。ところが興味深い研究が報告されている¹²⁾。原核生物であるグ

ラム陰性菌 *Klebsiella pneumoniae* (FG9株) の酵素であるブルナーゼのアミノ酸シーケンスが調べられた。Gly-X-Pro のリピードが6回くり返されているドメインが発見された。この酵素は, ホモトリマーを形成している。トリマー形成にこのコラーゲンドメインが必須であることが, コラーゲナーゼが消化実験からわかっている。またこの菌をコラーゲナーゼで消化するとブルナーゼが遊離することも調べられている。

10. おわりに

多細胞動物体制とコラーゲン分子の密接な関係は, コラーゲン遺伝子の構造解析が可能になってますますはっきりしてきた。しかしながら, 無脊椎動物のコラーゲン遺伝子と脊椎動物の25個のコラーゲン遺伝子の関係については, まだほとんど知られていない。それぞれのコラーゲン遺伝子の系譜が確立した時, 多細胞動物体制の進化が分子レベルで理解されるだろう。

またコラーゲン様の構造タンパク質の原始的な形が細菌類の酵素にみられることは, 留意しておく必要があるだろう。現代分子生物学の手法で, 単細胞動物におけるコラーゲン遺伝子の存否が調べられる日が近いことを期待しつつ拙稿を閉じたい。

文 献

- 1) Vuorio, E. & de Crombrughe, B.: *Annu. Rev. Biochem.*, 59, 837-872 (1990)
- 2) 林 利彦: 細胞外マトリックスのバイオサイエンスとバイオテクノロジー, p5-47, アイピーシー (1990)
- 3) Cox, G.N., Fields, C., Kramer, J.M., Rosenzweig, B. & Hirsh, D.: *Gene*, 76, 331-344 (1989)
- 4) Ninomiya, Y. & Olsen, B.R.: *Proc. Natl. Acad. Sci. USA*, 81, 3014-3018 (1984)
- 5) Von Mende, N., Bird, D.M., Albert, P.S. & Riddle, D.L.: *Cell*, 55, 567-576 (1988)
- 6) Kramer, J.M., Johnson, J.J., Edgar, R.S., Basch, C. & Roberts, S.: *Cell*, 55, 555-565 (1988)
- 7) Exposito, J-Y. & Garrone, R.: *Proc. Natl. Acad. Sci. USA.*, 87, 6669-6673 (1990)
- 8) Exposito, J-Y., Ouazana, R. & Garrone, R.: *Eur. J. Biochem.* 190, 401-406 (1990)
- 9) Shimizu, K., Amemiya, S. & Yoshizato, K.: *Biochim. Biophys. Acta*, 1038, 39-46 (1990)
- 10) D'Alessio, M., Ramirez, F., Suzuki, H., Solursh, M. & Gambino, R.: *Proc. Natl. Acad. Sci. USA*, 86, 9303-9307 (1989)
- 11) Venkatesan, M., De Pablo, F., Vogeli, G. & Shimpson, R.T.: *Proc. Natl. Acad. Sci. USA*, 83, 3351-3355 (1986)
- 12) Charalambous, B.M., Keen, J-N. & McPherson, M.J.: *EMBO J.*, 7, 2903-2909 (1988)