# 前胸腺刺激活性を持つカイコインシュリン様脳ホルモン，ボンビキシンの遺伝子構造



名古屋大学理学部
生物学教室
岩見雅史

# 主論文

# 目次

# 1．要約

ボンビキシン bombyxin は カイコ Bombyx mori の脳神経分泌細胞で合成され，エリサン Samia cynthia ricini の除脳蛹に対して，エクジソンの合成と分泌を促し変態を起こさせる活性を持つペプチドホルモンである．本研究では二種類のボンビキシン遺伝子（A－1およびC－1）を，カイコ遺伝子ライブラリーより単離しその構造を決定し，既知のインシュリン族遺伝子と比較した．これら二種類のボンビキシン遺伝子とそれぞれ類似した遺伝子群が存在し，ファミリー（AファミリーおよびCファミリー）を形成している．また，いずれのボンビキシン遺伝子もインシュリン遺伝子と塩基配列およびそのコードするアミノ酸配列レベルで相同性をもっている．ボンビキシン遺伝子によりコードされるボンビキシン前駆体（プレプロボンビキシン）はいずれもN端から順に，シグナルペプチド，B鎖，Cペプチド，A鎖の構成を持ち，プレプロインシュリンと全く同じ構成であり，エンドペプチダーゼにより転写後修飾をうけるCペプチド両端の塩基性アミノ酸配列も保存されている．これらのことからボンビキシンはインシュリン族に属し，その祖先遺伝子はヒトをはじめとする哺乳類のインシュリン遺伝子のそれと同一であると考えられる．しかし，少なくとも1個のイントロンを持つインシュリン遺伝子と異なり，ボンビキシン遺伝子にはイントロンが存在せず，プロセスされた遺伝子の特徴を備えている．ボンビキシンの一次構造をもとに，インシュリンと比較することにより予想される三次構造から導かれる特色として，1）インシュリンと同様の球状の構造を形成する，2）インシュリンのレセプター結合部位およびエピトープ部位に相当するアミノ酸配列がボンビキシンでは他のインシュリン族ペプチドと大きく異なる，3）インシュリンは二量体を形成し，二量体が集まり六量体を形づくっているが，ボンビキ

シンには二量体を形成するためのアミノ酸残基が欠けていることから，ボンビキシンは一量体として存在する，の諸点があげられる．さらに，ボンビキシンCファミリー分子は，ボンビキシン内でも，他のファミリーと進化的および機能的に，大きくへだたっている．

## 2. 序論

前胸腺刺激脳ホルモン Prothoracicotropic hormone (ＰＴＴＨ) は，昆虫の脳に存在する神経分泌細胞で合成され，前胸腺に働き，ステロイドホルモンであるエクジソンの合成と分泌を促す (11,29). エクジソンは体内の諸器官に作用し，昆虫の成長や変態を引き起こす．すなわち，ＰＴＴＨはエクジソンの合成と放出を制御することにより，昆虫の成長や変態を最も上位で支配しているホルモンである．

カイコ Bombyx mori の脳内には，カイコ自身のＰＴＴＨ（分子量 約三万，前称 22Ｋ－ＰＴＴＨ）に加えて，ボンビキシン（分子量約五千，前称 4Ｋ－ＰＴＴＨ）が存在している．ボンビキシンは，カイコ（カイコガ科）自身に対しては変態を起こさないが，ヤママユガ科のガであるエリサン Samia cynthia ricini の除脳蛹に対して変態を引き起こす活性を持つ(28,30,31,68).

ボンビキシンは，それぞれＡ，Ｂ二つのペプチド鎖より成るヘテロダイマーで，わずかにアミノ酸配列の異なる多数の分子種が存在する．現在までに，ボンビキシンの全一次構造がボンビキシン－II (51-53)とボンビキシン－IV(44,53)について，また部分的な一次構造が ボンビキシン－I，－III，－V (35,51) について決定されている．これらボンビキシン分子は，アミノ酸レベルでインシュリン族分子と相同性を示している．さらに，Ａ，Ｂ二つのペプチド鎖はインシュリンと同様のジスルフィド結合で結ばれており(53)，コンピューターを使った構造解析は，ボンビキシンがインシュリンに似た球状の三次構造をとることを示している(35)．ボンビキシンのエリサンに対する明快な生理活性に比べ，カイコ自身に対する生理作用は現在不明であるが，インシュリン族ホルモンであることから重要な生理機能を担っていると考えられる．

私は以下に述べる目的のため，ボンビキシン遺伝子の単離を試みた．

1）ペプチド分析では得られないボンビキシン前駆体の構造，および発現調節に直接関わりを持つプロモーター等の構造をはじめとする遺伝子解析は，ボンビキシンのみならず昆虫ホルモン全般に対して貴重な知見を与える．

2）免疫学的，生物学的方法によって，無脊椎動物多種にわたるインシュリン様分子の存在が示唆されており(40,42,70)，その構造を決定することはインシュリン分子の進化を考える上で重要である．さらに，その構造をヒトなどの脊椎動物のインシュリンと比較することにより分子構造と機能との関係を議論するための材料を提供することができる．

2）単離された遺伝子の分子プローブとしての使用は，ボンビキシンが持つと思われる重要な生理作用を探るために有効な手段を提供しうる．

3）ペプチド分析により，ボンビキシンは構造のわずかに異なる数多くの分子種からなることが示されている(51-53)．この構造の違いは，ボンビキシンをコードする数多くの遺伝子の存在によるものか，スプライシング等転写後の修飾によるものか，もしくは，カイコの品種の違いによるものか，ボンビキシン遺伝子を単離することで明らかにすることができる．現在までに調べられた限り，インシュリンもしくはインシュリン様分子をコードする遺伝子はゲノム当り1コピーか2コピーである(66)．もしカイコにおいて，インシュリン様分子をコードする遺伝子が数多く存在するなら興味深い知見となる．

本論文では，互いに異なる二種類のボンビキシン遺伝子（A－1およびC－1）の単離と構造解析について述べ，さらに他のインシュリンおよびインシュリン様分子の遺伝子構造との比較考察について論議する．

# ３．材料と方法

## ３－１　オリゴヌクレオチドプローブ

ボンビキシン－ＩＩのアミノ酸配列(52)をもとに，二種類の合成オリゴヌクレオチドプローブを準備した（図１）．プローブＡは，Ａ鎖のＮ端から4番目のAspから20番目のCysまでの 17個のアミノ酸に対する51塩基からなるオリゴヌクレオチドで，カイコ体液タンパク質30K(59)で最も頻繁に使われているコドンを使用して合成したものである．プローブＢは，Ｂ鎖のＮ端から8番目のThrから24番目のGluまでの 17個のアミノ酸に対する51塩基からなるオリゴヌクレオチドで，ショウジョウバエ(Drosophila melanogaster)の コドン利用度(43)を参考に合成したものである．

## ３－２　ゲノムライブラリーの作成

カイコＤＮＡ（錦秋 X 鐘和；絹糸腺由来）を Sau3AIで部分消化し，約10-23キロ塩基対のＤＮＡ断片を得る．このＤＮＡ断片を ウシ腸由来アルカリフォスファターゼで脱リン酸化処理した後，λEMBL ＤＮＡ(24)を BamHIとEcoRIで完全消化し作成した BamHIアームにＴ4ＤＮＡリガーゼで接続し，in vitro パッケージング反応を行なった．36万個の独立した組換え体ファージからなるゲノムライブラリーが得られた．

## ３－３　ボンビキシン遺伝子の単離

ボンビキシンＡ－ファミリー遺伝子の単離には，12万クローンより成るゲノムＤＮＡライブラリーに対し，まず［$\gamma-^{32}P$］ＡＴＰとＴ4ポリヌクレオチドキナーゼにより 5'末端をラベルしたプローブＡで，プラークハイブリダイゼーション法(45)によりスクリーニングを行なった．オリゴヌクレオチドプローブ（プローブＡおよびＢ）とのハイブリダイゼーションおよび洗いの条件は，Latheの式

(41):

$$T = 94-(84/L)-1.2(100-H)$$

T: 2X ＳＳＣ下でのフィルターの洗いの温度

L: プローブの長さ（塩基数）

H: プローブと目的のＤＮＡとのホモロジー（％）

ただし，ＳＳＣ: 0.15M Nacl, 0.015M Na-citrate.

を使い，Ｌ＝51，Ｈ＝78　として，Ｔ＝52　を導いた．

ハイブリダイゼーションは，6X ＳＳＣ，10X Denhardt(Denhardt：0.02% ウシ血清アルブミン，0.02% Ficoll，0.02% polyvinyl-pyrrolidone)，1% ＳＤＳ，0.5mg/ml 酵母ｔＲＮＡ，1pmol/ml プローブＤＮＡを含む溶液中，52°Ｃで20時間行なった．洗いは，52°Ｃで，6X ＳＳＣ，0.1% ＳＤＳ下で 2時間，続いて，2X ＳＳＣ，0.1% ＳＤＳ 条件下で30分間行なった．増感スクリーンを用いたオートラジオグラフィーにより，目的のファージを検出した．スクリーニングは単一の集団からなる組換え体ファージが得られるまで少なくとも3回繰り返した．

ボンビキシンＣ－ファミリー遺伝子の単離には，λＢｂ２０４（ボンビキシンＣ－ファミリー遺伝子の3'側一部領域を含むｃＤＮＡクローン；１参照）をプローブとして用い，12万クローンより成るゲノムＤＮＡライブラリーに対して，スクリーニングをプラークハイブリダイゼーション法(45)により行なった．

ハイブリダイゼーションは，6X ＳＳＣ，10X Denhardt，1% ＳＤＳ，10% Dextran sulfate，0.25mg/ml サケ精子ＤＮＡ中で，ランダムヘキサヌクレオチドと［$\alpha-^{32}P$］ｄＣＴＰでラベルした(21)プローブと，65°Ｃで 20時間 反応させた．洗いは，65°Ｃで，2X ＳＳＣ，0.1% ＳＤＳ 下で 2時間，引き続き，0.2X ＳＳＣ，0.1% ＳＤＳ下で３０分間行なった．オリゴヌクレオチドプローブによるスクリーニングと同様に，オートラジオグラフィーにより，目的のファージを検出した．スクリーニングは単一の集団からなるファージが得られるまで少なくとも３回繰り返した．

３－４　　ボンビキシン遺伝子の解析

オリゴヌクレオチドプローブＡを用いてのスクリーニングにより得られたクローン６個（λ４Ｋ１０１，λ４Ｋ１０５，λ４Ｋ１０７，λ４Ｋ１０８，λ４Ｋ１１１，λ４Ｋ１１２）のうち，λ４Ｋ１０８とλ４Ｋ１１１を除く４個のクローンは，プローブＢに対しても陽性であった（図２）．これら４個のクローンのうち，ボンビキシン遺伝子を１コピー持つλ４Ｋ１１２にしぼり解析を行なった．

プローブＡ，Ｂいずれに対しても陽性の HindIII断片（約１キロ塩基対）を pUCベクター(77)にサブクローニングし，エキソヌクレアーゼ－III を用いて(25)，いくつかのデレーションミュータントを作成した．これらミュータントを利用し，塩基配列の決定を行なった．塩基配列決定の方式を図３ａに示した．また，これらミュータントのいくつかをＤＮＡプローブとして用いた．塩基配列は，ジデオキシヌクレオチド法(60)を改良し，7-deaza－ｄＧＴＰと［α－$^{35}$Ｓ］ｄＣＴＰをつかう方法(49)により決定した．塩基配列は両方の鎖について決定した．

λＢｂ２０４プローブを用いたスクリーニングにより５個のクローン（λ４Ｋ４０１，λ４Ｋ４０２，λ４Ｋ４０３，λ４Ｋ４０５，λ４Ｋ４０６）が得られた．これらクローンのうち，λ４Ｋ４０３にしぼり解析を行なった．プローブに対して陽性であるいくつかの制限酵素断片を pUCベクターにサブクローニングし，図３ｂに示した戦略で塩基配列を決定した．塩基配列決定法は上の場合と同様である．得られた塩基配列は，いずれも塩基配列解析ソフトウエア（ＤＮＡＳＩＳ，日立ソフトウエア）により解析した．

３－５　　遺伝子間の進化的距離

二つの遺伝子間のヌクレオチド座位当りの進化的距離（Ｋ）は木村(39)の３ＳＴ（three-substitution-type）法により求めた．すなわち，塩基転位型変換（Ｕ－Ｃ，Ａ－Ｇ）の頻度をＰ，塩基転換型

変換（Ｕ－Ａ，Ｃ－Ｇ）の頻度をQ，塩基転換型変換（Ｕ－Ｇ，Ｃ－Ａ）の頻度をＲとすると，

$$K = -(1/4)\ln[(1-2P-2Q)(1-2P-2R)(1-2Q-2R)]$$

の式によりＫの値を求めた．

３－６　ノーザンハイブリダイゼーション

ＲＮＡフィルター［絹糸腺ポリ(Ａ)ＲＮＡ(5μg)と４－５令幼虫脳ポリ(Ａ)ＲＮＡ(3.3μg)を転写したもの，安達（名古屋大学）より］に，［α－$^{32}$Ｐ］ｄＣＴＰで それぞれニックトランスレーションしたプローブａ（ｐ４ＫＧＥＭ２０;図４の塩基配列505から1045までを含む）またはプローブｂ（ｐ４ＫＨＤ３４;図４の塩基配列1から564までを含む）を，50% formamide, 5X ＳＳＣ，1% ＳＤＳ，10% Dextran sulfate, 0.3mg/ml サケ精子ＤＮＡ溶液中で，42°Ｃ 16時間ハイブリダイゼーションさせた．洗いは，65°Ｃで，2XＳＳＣ，0.1% ＳＤＳ条件下2時間，続いて，0.2X ＳＳＣ，0.1% ＳＤＳ条件下30分間行なった．増感スクリーンを用い，オートラジオグラフィーにより，シグナルを検出した．

３－７　サザンハイブリダイゼーション

カイコゲノムＤＮＡをBamHIまたはHindIIIで完全消化した後，アガロースゲル（0.7%）電気泳動で分画し，ナイロンメンブレンに，0.4N NaOH で転写させた(56)．ニックトランスレーションしたプローブａ（ｐ４ＫＧＥＭ２０）を，6X ＳＳＣ，10X Denhardt, 1% ＳＤＳ，10% Dextran sulfate, 0.3mg/ml サケ精子ＤＮＡ溶液中，65°Ｃ 20時間ハイブリダイゼーションさせた．洗いは，ノーザンハイブリダイゼーションと同一条件下で行なった．増感スクリーンを用い，オートラジオグラフィーによりシグナルを検出した．

# ４．結果

## ４－１　ボンビキシンＡ－１遺伝子の構造

オリゴヌクレオチドプローブA，Bのいずれに対しても陽性であるλ４Ｋ１１２由来の HindIII断片（1045塩基対）の塩基配列を図４に示した(32)．このＤＮＡ断片中には，ボンビキシンのＡ鎖20アミノ酸残基と，Ｂ鎖28アミノ酸残基をコードしうる 276塩基対からなるオープンリーディングフレームが存在している．オープンリーディングフレームの上流域には，CAT-box (GGCAATGT)および TATA-box(TATATAA)様塩基配列が，下流域にはポリ(Ａ)付加シグナル(AATAAA)，12塩基対の保存配列（ATACAGTTTTTT; 56参照）が存在しており，このオープンリーディングフレームは ＤＮＡポリメラーゼ-IIにより転写されうる遺伝子と考えられる．

Ａ鎖20アミノ酸残基，Ｂ鎖28アミノ酸残基のうち，Ｂ鎖５番目のアミノ酸残基Argを除き ボンビキシンーIIのアミノ酸配列(52)と一致した．ペプチド分析から，ボンビキシンには一次構造の異なる分子が数多く存在することが示されている．しかも，Ｂ鎖５番目のアミノ酸残基は変異が比較的集中している箇所である．以上のことから，このオープンリーディングフレームがコードするペプチドは，ボンビキシンーIIのバリアントと考えられる．この遺伝子は，4-2項「ボンビキシンＡ－１遺伝子の発現とコピー数」で述べるように，構造の近似した複数の遺伝子群と共に一つのファミリーを形成していることから，ボンビキシンＡ－１（Ａ－ファミリーの1番目の意）と名付けた．ボンビキシンーIIＢ鎖Ｎ端の pyroglutamateは，翻訳後の修飾によりGlnより 生合成されたと考えられる(72)．

翻訳開始部位をコードしうる塩基配列が，-5，-10，-19に存在している．シグナルペプチドの一般的特徴(75)およびショウジョウバエの翻訳開始部位の保存配列(13)から，Met -19 が翻訳開始点と考えられる．Ｂ鎖とＡ鎖との間には，25アミノ酸残基のペプチド（Ｃ

ペプチド）が介在し，Ｃペプチドの両端は 塩基性アミノ酸残基対(Lys-Arg)で構成されている．この塩基性アミノ酸残基対は，トリプシン様エンドペプチダーゼによる切断認識部位と考えられ(14)，インシュリンの Ｃペプチドでも保存されている(3)．このことは，ボンビキシンＡ－１も インシュリンと同様な翻訳後の修飾を受け(3, 14)，プレプロボンビキシンからＣペプチドが除去されて，Ａ，Ｂ鎖から成る成熟ボンビキシンが生合成されることを示唆している（図５）．

ボンビキシンＡ－１では，Ａ鎖に続いて停止コドン（アンバー；TAG）があり，インシュリン様成長因子（ＩＧＦ）－I，－II(5,34)と違いＤペプチドやＥペプチドをコードしていない．プレプロボンビキシン遺伝子の構成は，5'側から，シグナルペプチド，Ｂ鎖，Ｃペプチド，Ａ鎖とコードするものであり，インシュリン遺伝子(4)と全く同じである．ただし，現在までに調べられたすべてのインシュリン遺伝子が，少なくとも１箇所，イントロンを持っているのに対し(66)，ボンビキシンＡ－１遺伝子にはイントロンが存在しない．また，ボンビキシンＡ－１遺伝子の転写領域の両端付近には，転写領域をはさむように，６塩基対(TTCTTC)の反復配列が存在している．さらに，3'側の反復配列の左側に，ポリ(Ａ)配列の残存と考えられうる配列が存在していた（図４波線部）．これらのことは，ボンビキシンＡ－１遺伝子がプロセスされた遺伝子であることを示唆している (32, 33)．

４－２　　ボンビキシンＡ－１遺伝子の発現とコピー数

ボンビキシンはカイコの脳に存在する４対の神経分泌細胞でつくられ，アラタ体を経て体液中に放出される(48)．ボンビキシンＡ－１遺伝子をプローブとして検出可能な転写産物がカイコ脳中に存在するか，ノーザンハイブリダイゼーションにより調べた（図６ａ，ｂ）．ボンビキシンＡ－１遺伝子より上流域（ノンコーディング領域）をプローブとした場合は，絹糸腺ポリ(Ａ)ＲＮＡ，４－５令幼

虫脳ポリ(A)ＲＮＡのいずれにもシグナルが検出できないが，Ａ－１遺伝子のコーディング領域をプローブとした場合は，脳に約600塩基の長さに相当するｍＲＮＡが検出できた．この長さは，Ａ－１遺伝子より転写されるであろうｍＲＮＡの長さによく一致している．

ボンビキシンＡ－１遺伝子をプローブに用いたサザンハイブリダイゼーションを行なった（図６ｃ）．HindIIIで消化した カイコゲノムＤＮＡでは，約１キロから20キロ塩基対にかけて数本以上のシグナルが検出できる．しかも，約１キロのシグナルは，ボンビキシンＡ－１遺伝子をコードしており，シグナルの濃さはボンビキシンＡ－１遺伝子１コピーに相当する．このことから，Ａ－１遺伝子で検出可能な遺伝子，すなわちＡ－ファミリー遺伝子は，カイコゲノム中に，10コピー以上 存在していると考えられる．また，BamHIで消化したＤＮＡでは，約10キロから30キロ塩基対領域に３本のシグナルが検出できる．すなわち，ボンビキシンＡ－ファミリー遺伝子は，カイコゲノムの3-4箇所に 遺伝子群を形成していると推定できる．ボンビキシンＡ－ファミリーとは異なる別のボンビキシンファミリーであるＢ－ファミリーの遺伝子も，カイコゲノム中に少なくとも10コピー存在することが示されている(38)．

４－３　　ボンビキシンＣ－１遺伝子の構造

λＢｂ２０４プローブに対して陽性であるλ４Ｋ４０３由来のHindIII/KpnI断片（456塩基対）の塩基配列を図７に示した． この断片中に存在する273塩基対からなる オープンリーディングフレームは，ボンビキシンＡ－１遺伝子，およびＢファミリー遺伝子(38)とあるていど塩基配列の相同性をもっているが，明らかに違ったペプチドをコードしている．しかも，このペプチドのアミノ酸配列は，現在までに一次構造の全部もしくは一部が決定されたボンビキシン－I，－II，－III，－IV，－V(34,44,51-53)のいずれとも一致しない．しかも，この遺伝子もカイコゲノム当り複数個存在する（投稿中）ことから，この遺伝子をボンビキシンＣ－１遺伝子（Ｃ－ファ

ミリーの１番目の意）と名付けた．ボンビキシンＣ－１のＢ鎖Ｎ端のGlnは，ボンビキシンＡ－１と同じように，翻訳後の修飾によりpyroglutamateに転換されると考えられる．また，ボンビキシンＣ－１をコードするオープンリーディングフレームの下流域には，ポリ(Ａ)付加シグナル(AATAAA)が存在する．

翻訳開始部位をコードしうる塩基配列が，-7，-15，-19に存在している．シグナルペプチドの一般的特徴(75)，およびショウジョウバエの翻訳開始部位の保存配列(13)から，ボンビキシンＡ－１と同様，Met -19 が翻訳開始点と考えられる．Ｂ鎖Ｃ端とＡ鎖Ｎ端との間には，Ｂ鎖とＡ鎖を結びつける25アミノ酸残基のＣペプチドが存在し，Ｃペプチドの両端は 塩基性アミノ酸残基対(Lys-Arg)で構成されている．この塩基性アミノ酸残基対は，インシュリンのＣペプチドでも保存されている(3)．このことは，ボンビキシンＣ－１もインシュリンやボンビキシンＡ－１と同様な翻訳後の修飾を受け，プレプロボンビキシンから最終的なボンビキシンが生合成されることを示唆している．

ボンビキシンＣ－１遺伝子においても，Ａ鎖に続いてアンバー停止コドンがある．プレプロボンビキシンＣ－１遺伝子の構成は，5'側から，シグナルペプチド，Ｂ鎖，Ｃペプチド，Ａ鎖であり，インシュリン遺伝子(4)，ボンビキシンＡ－１遺伝子と全く同じである．さらに，ボンビキシンＣ－１遺伝子にも，Ａ－１と同様に，イントロンが存在せず，ボンビキシンＣ－１遺伝子の転写領域の両端付近にも，転写領域をはさむように，不完全ではあるが反復配列（5'側，AATATT；3'側，AATATAT）が 存在している．これらのことは，ボンビキシンＣ－１遺伝子もプロセスされた遺伝子であることを示唆している(33)．

## ４－４　　ボンビキシン遺伝子間の塩基配列の相同性

ボンビキシンＡ－１遺伝子を，ボンビキシンＢファミリー遺伝子の一つのＢ－１遺伝子(38)と，ハープロット解析により塩基配列を

比較した（図８ａ）．Ａ－１，Ｂ－１間で，Ｂ鎖，Ｃペプチド，Ａ鎖において塩基配列の高い相同性がみられる．しかし，相同性の度合は，Ｂ鎖ではそれほど高くない．さらに，Ｃ－１とＡ－１間，Ｃ－１とＢ－１間では，Ｂ鎖で塩基配列の相同性はあまりみられず，ＣペプチドとＡ鎖のみに 高い相同性がみられる（図８ｂ，ｃ）．インシュリン族遺伝子では，Ａ鎖およびＢ鎖での相同性が種や遺伝子の違いを越えてよく保存されているが(66)，ボンビキシン遺伝子間では，いずれもインシュリン族遺伝子でありながら，塩基配列は，Ｂ鎖ではあまり保存されておらず，ＣペプチドとＡ鎖においてよく保存されている．

このことは，木村の式(39)による，ボンビキシン遺伝子間のヌクレオチド座位当りの進化的距離（表１）においてもよく表されている．すなわち，ボンビキシンＡ－１とＢ－１遺伝子間の進化的距離は，Ａ鎖，Ｂ鎖，Ｃペプチド共に，0.5 前後である．これに対し，Ａ－１とＢ－１間，Ｂ－１とＣ－１間においては，Ａ鎖，Ｃペプチドでは，いずれも0.5前後であり，Ｂ鎖では順に 0.75，1.02と値が大きくなっている．このＢ鎖におけるＣ－１遺伝子と他のボンビキシン遺伝子との進化的距離の大きさは，従来のインシュリン族遺伝子のＢ鎖における進化的距離を，Ａ鎖およびＣペプチドにおける進化的距離と比較した際の値と大きく異なっている（たとえば表１最下段のヒト，ラットインシュリン間の進化的距離）．ボンビキシンＣ－１遺伝子Ｂ鎖で塩基の置換率が高いことは，Ａ－１，Ｂ－１のＢ鎖およびＡ－１，Ｂ－１，Ｃ－１のＡ鎖とＣペプチドが 機能的必要性から 置換率が低いのに対し，Ｃ－１のＢ鎖は他ファミリーのＢ鎖が持つ共通の機能をになう制約からのがれたために，置換率が高くなっていると考えられる．このことは，ボンビキシンＣファミリーが機能的に，特にＢ鎖の機能において，他のＡ，Ｂファミリーと隔たりがあることを示唆している（5-2 項「予想されるボンビキシンの三次構造及び生理機能」で詳述）．実際，ボンビキシンＣファミリーのｍＲＮＡは，卵からさなぎ期まで発現しているのに対し，

AおよびBファミリーmRNAは，2令もしくは3令期以降しか発現していない（川上ら未発表）．

4－5　　ボンビキシン間のアミノ酸配列の比較

ボンビキシン遺伝子から予想されるプレプロボンビキシンのアミノ酸配列をボンビキシン天然標品から得られたアミノ酸配列，ヒトプレプロインシュリン，モルモットプレプロインシュリン，IGF－I，－II，および ヒトリラキシンのアミノ酸配列と比較したのが図9である．A鎖において，ボンビキシンAファミリーはBファミリーと4残基，C－1はAファミリーと4残基，Bファミリーと5残基に違いがみられる．B鎖については，A，B両ファミリー間でアミノ酸配列の相同性が 61%であるのに対し，A，C間，B，C間ではそれぞれ 44%，36%しかない．特筆すべきは，B9，B10，B16，B24の4残基が，A，Bファミリー内では 完全に保存されているにも拘らずC－1ではすべて置換されていることである．この結果は塩基配列レベルでの結果と同じく，CファミリーがA，Bファミリーと明らかに違ったグループに属することを裏付けている．

4－6　　ボンビキシンとインシュリン両遺伝子間の塩基配列及びアミノ酸配列の比較

ボンビキシンA，B，C各ファミリー遺伝子の代表として，A－1，B－1，C－1遺伝子を，ヒトインシュリン cDNA(4)と，ハープロット解析により，塩基配列を比較した（図10）．A－1とインシュリン間，B－1とインシュリン間，C－1とインシュリン間のいずれにおいても，A鎖の一部を除いて，塩基配列の相同性は低い．このことから，ボンビキシンとインシュリンは同じインシュリン族に属しながら，進化的に隔たったものであり，別のグループに属すると考えられる．

しかし，A鎖に関しては，ボンビキシン各ファミリーとインシュリン間で比較的塩基配列が保存されていることから，構造的もしく

は機能的に重要な役割を果たしていると推定される.

塩基配列が比較的保存されているA鎖について，ボンビキシンとヒトインシュリン族ペプチド間でアミノ酸配列の相同性を百分率で示したのが表2である．ボンビキシンはアミノ酸レベルにおいてもインシュリン族ペプチドと高い相同性を示している．注目に値するのは，ヒトインシュリンとヒトＩＧＦ－Ｉ(5)間の相同性が，55%を示しているのに対し，ボンビキシンとＩＧＦ－Ｉ間ではさらに高い値（Ａ－１，60%；Ｂ－１，65%；Ｃ－１，60%）を示していることである.

# 5 . 議論

## 5－1　インシュリン族としてのボンビキシン遺伝子

二種類のボンビキシン遺伝子（A－1およびC－1）をカイコゲノムより単離し，塩基配列決定を含む構造解析を行なった．これらボンビキシン遺伝子は，昆虫ペプチドホルモンで最初に単離され，構造の決定された遺伝子である．いずれのボンビキシン遺伝子もインシュリン遺伝子と塩基配列およびアミノ酸配列レベルで相同性をもっている．ボンビキシン遺伝子によりコードされるボンビキシン前駆体（プレプロボンビキシン）は，いずれもN端から順にシグナルペプチド，B鎖，Cペプチド，A鎖の構成をもち，インシュリンと全く同じで，さらにエンドペプチダーゼにより転写後修飾をうけるアミノ酸配列も保存されている．塩基配列の相同性，前駆体の遺伝子構成，転写後修飾を受ける部位の構造的特色から，ボンビキシンがインシュリン族に属することが示された（図11）．これら事実は，昆虫の脳ホルモンであるボンビキシンが，脊椎動物の消化管ホルモンであるインシュリンと共通の祖先を持つことを強く示唆している．

## 5－2　予想されるボンビキシンの三次構造および生理機能

ボンビキシンの一次構造をもとに，すでにX線結晶解析により三次構造が解明されたインシュリンと比較することにより予想されるボンビキシンの三次構造から導かれる特色，およびペプチドホルモン一般の一次構造から予想される特色として次のことがあげられる．

### 5－2－1　シグナルペプチド

シグナルペプチドは，翻訳されつつあるペプチドを小胞体へ導くシグナルを有する 約20-60アミノ酸残基からなる短いペプチドである．シグナルペプチドはアミノ酸配列の特徴から三つのサブドメイ

ンに分けることができる(74). 一つは N端から5個程度のアミノ酸残基で，このサブドメイン中に電荷を帯びたアミノ酸残基（Lys等）が含まれている．次の一つは，上記5個程度に続いて少なくとも9個のアミノ酸残基からなる疎水性アミノ酸で構成されており膜を貫通しうるサブドメインである．最後のサブドメインは，C端に近い4-8個のアミノ酸残基からなり，ヘリックス構造を壊しうる アミノ酸残基（GlyもしくはPro），または極性の強いアミノ酸残基（Glu等）で構成されているサブドメインである．

また，翻訳後の修飾（切断）を受ける部位には，小さな，電荷を持たない側鎖を持つ アミノ酸残基(Ala, Gly, Ser, Cys, Thr)を持つことが知られている(74).

ボンビキシンのシグナルペプチドは，各ファミリー間でアミノ酸配列が，S2 Lys（インシュリンを基準にして，シグナルペプチド2番目の位置に存在するLysの意），S8 Leu，S16 Leu以外保存されていない（図９）．また，インシュリン族ペプチドのシグナルペプチドについても同様に，各ペプチド間でアミノ酸配列は，S3 Leu，S11 Leu，S12 Ala，S13 Leu，S25 Ala，S16 Leuを除いて保存されていない．しかし，インシュリン族ペプチドのシグナルペプチドはむろん，ボンビキシンのシグナルペプチドも，アミノ酸配列の相同性はないものの，シグナルペプチドのもつ一般的特徴を満たしている．さらに，翻訳後の修飾（切断）を受ける部位のアミノ酸残基（ボンビキシンA－1:Thr，B－1:Ala，C－1:Ala）も上に述べた特徴を満たしている．

５－２－２　Ｂ鎖

ボンビキシン，インシュリン族ペプチドのすべてのＢ鎖で，三次構造の疎水性の核を形成するためのジスルフィド結合を与えるアミノ酸残基B7 Cys，B19 Cys，三次構造形成を促すB8 Glyが保存されている（図９）．しかし，ボンビキシンにはインシュリンとアミノ酸配列の異なるサブドメインが３箇所存在する．

インシュリンでは，Ｂ鎖Ｎ端付近のアミノ酸残基（Ｂ1－Ｂ4）は，Ａ鎖アミノ酸残基Ａ8 Thr，Ａ9 Ser，Ａ10 Ileと共に，親水性のドメインを形成し，抗インシュリン抗体の認識部位となっている（7，10）．このサブドメインの アミノ酸配列は，ボンビキシンとインシュリン族ペプチド間ではむろん，ボンビキシン各ファミリー間でも異なっている．このことは，抗インシュリン抗体がボンビキシンに対して反応しないであろうこと，また，ボンビキシン各ファミリーに対して特異的な抗体の作成が可能であろうことを示している．抗ブタインシュリン抗体がボンビキシン－IIと反応しない事実は(51)，上の仮説を支持している．

インシュリンＢ鎖Ｂ12 SerからＢ22 Cysにかけてのサブドメインは，インシュリン分子の疎水性の核を構成するα－ヘリックスを形成する(8,76)．ボンビキシンのこのサブドメインにおけるアミノ酸配列は，インシュリンと同様，疎水性が強く，疎水性のドメインを形成し，インシュリン様の球状の核を形づくっていると考えられる．

インシュリンＢ鎖Ｂ24 Phe，Ｂ25 Phe，Ｂ26 Tyr，Ｂ27 Thrのサブドメインは，Ｂ16，Ａ鎖Ｎ端部と共に，二量体を形成する水素結合の供給源，さらには，インシュリンがインシュリンレセプターと結合する際の結合部位となっている(7-9,46,55,76)．インシュリンのこのサブドメインにおけるアミノ酸配列のＩＧＦ－I，－II に対する類似（とくにＢ24 Phe，Ｂ26 Tyr）により，インシュリンはインシュリンレセプターと 結合するのに加えて，ＩＧＦ－I およびＩＧＦ－IIレセプターとも結合することが示されている(37)．ところが，ボンビキシンは，このサブドメインにおけるアミノ酸配列がインシュリン，ＩＧＦのいずれとも大きく異なっている．このことは，ボンビキシンはインシュリンレセプターと結合しない，すなわち，ボンビキシンレセプターとインシュリンレセプターとは異なった分子であることを示唆している．事実，ショウジョウバエのインシュリンレセプターは，進化的に隔たっているブタのインシュリンと結合するが，より進化的に近いカイコのインシュリン族ペプチド

であるボンビキシン（ボンビキシンーII）とは結合しない(22)．さらに，このサブドメインでは，ボンビキシン間でもアミノ酸配列が異なっている．すなわちC－1では，インシュリンにおけるレセプター結合部位である B16，B24，そしてその近傍に位置するB9，B10がボンビキシンAおよびBファミリーと異なることにより（A，Bファミリー間ではこれらアミノ酸配列は完全に保存されている），ボンビキシンAおよびBファミリーとCファミリーに対する別々のレセプターが存在する可能性があることをも示唆している．実際ペプチド分析からは，エリサン前胸腺刺激活性をただ一つの指標としたボンビキシンの精製，構造決定において，Cファミリーに相当する分子は見つかっていない．

モルモットなどのヤマアラシ類をのぞくインシュリンには，B10の位置に Hisが存在する．この保存された His残基はインシュリンが六量体を形成する際，亜鉛と結合するために不可欠なアミノ酸残基である(7,10)．ボンビキシンは二量体も六量体も形成しないと考えられることから，A，Bファミリー B10 に存在するHisは 二次的類似によるものと予想されていた(35)．ボンビキシンC－1でB10 にPheが存在することは，この予想を正当化するものであろう．

### 5－2－3　Cペプチド

Cペプチドは，A鎖とB鎖が形づくる球状の三次構造を，効率よく形成させるためのペプチドと考えられている(65)．そのために，インシュリンのCペプチドは，適当な長さと電荷を持っている（14,15,65,66)．さらに，高等な脊椎動物では，Cペプチドの中央部にβ－ベンドを形成するGlyに富んだ部分がある(14，15,66)．

ボンビキシン各ファミリー間のCペプチドのアミノ酸配列は，保存置換を含めるとよく保存されている（図９）．各ファミリーのCペプチドが互いに似た三次構造をとり，類似の機能を有すると考えられる．しかし，ボンビキシンとインシュリン間において，Cペプチドのアミノ酸配列の相同性はない．このことは，Cペプチドも，

シグナルペプチドと同様に，個々のアミノ酸配列より極性や電荷などの要素を満たすことが機能発現に重要と考えられる．

ボンビキシンのＣペプチドは，翻訳後の修飾を受けないＩＧＦ－Ｉ，－ＩＩ のＣペプチドとは長さおよびアミノ酸配列が異り，ＩＧＦ族とはペプチド構成上からは，すなわち翻訳後の修飾を含む生合成経路からは，別のグループに属すると考えられる．

ボンビキシンＣペプチドの両端に存在する塩基性アミノ酸残基対は，ボンビキシンもインシュリンと同様なトリプシン様エンドペプチダーゼにより，翻訳後の修飾（Ｃペプチドの除去）が行なわれることを予想させる．

５－２－４　Ａ鎖

ボンビキシン各ファミリーでＡ鎖の長さが異なる．Ｂファミリーはインシュリンと同じで21アミノ酸残基，Ａファミリーはヒトリラキシンと同じで20アミノ酸残基，Ｃファミリーは22アミノ酸残基より成る．これらの長さの差は，Ｂ20 CysよりＣ端側のアミノ酸残基の延長に起因する（図９）．

ボンビキシンＡ鎖は，各ファミリー間で，アミノ酸配列がよく保存されている．ボンビキシンＡ鎖とインシュリンＡ鎖を比べた場合，アミノ酸配列の相同性は，保存置換も含めると，高い．インシュリン族ペプチドＢ鎖と同様に，三次構造の疎水性の核を形成するためのジスルフィド結合を与えるアミノ酸残基（Ａ6 Cys，Ａ7 Cys，Ａ11 Cys，Ａ20 Cys），三次構造の形成を促すアミノ酸残基（Ａ16 Leu，Ａ19 Tyr）は，すべてのインシュリン族ペプチドの間で 保存されている．しかし，アミノ酸配列が保存されていないサブドメイン(Ａ8－Ａ10)が 存在している．このサブドメインは，さきに述べた通り，インシュリンとの比較から，Ｂ鎖Ｎ端部位と共に親水性のドメインを形成し，ボンビキシンに対する抗体の認識部位であると考えられる．

５－３　　ボンビキシン遺伝子のゲノム当りのコピー数

既知インシュリン族遺伝子はゲノム当り1もしくは2コピー存在している(66)．ところが，カイコゲノム中にはボンビキシン遺伝子が少なくとも20コピー存在している．このカイコにおけるインシュリン族分子をコードする遺伝子のコピー数の多さは，カイコが発生のある段階において，大量のボンビキシンを必要とするためかも知れない．大量の転写産物をまかなうため，昆虫はある発生時期に遺伝子を増幅させる方法［ショウジョウバエのコリオン遺伝子(36)など］と，最初から重複した遺伝子をゲノム中に持つ方法［カイコのコリオン遺伝子(36)など］で対処する場合がある．カイコのボンビキシンの場合は，後者で対処していると考えられる．

５－４　　ボンビキシン遺伝子はプロセスされた遺伝子である

ボンビキシン遺伝子は脊椎動物のインシュリン族遺伝子といくつかの点で異なっている．まず，現在までに構造の決定されたすべてのインシュリン族遺伝子が，少なくとも一つイントロンを持っているのに対し(66)，ボンビキシン遺伝子はイントロンを持っていない．さらに，ボンビキシン遺伝子には，転写されるであろう領域の前後に，6塩基対もしくは7塩基対の反復配列が存在している（図１２）．さらに，Ａファミリー遺伝子には，ポリ(Ａ)配列の残存配列が存在している．これら三つの特色は，スプライシングを受けたｍＲＮＡが逆転写され再びゲノムに挿入されたと考えられるプロセスされた遺伝子，もしくは レトロポゾン(58)のもつ特徴(61,73)であり，ボンビキシン遺伝子はプロセスされた遺伝子と考えられる (32,33)．しかし，ボンビキシンＡ－１遺伝子には，ＲＮＡポリメラーゼ-IIで転写されうるプロモーター，また，ポリ(Ａ)付加シグナル等が備わっており機能しうる遺伝子と予想される．さらに，ボンビキシンＣ－１遺伝子の 3'側約半分の塩基配列は 脳のｃＤＮＡライブラリーより得られたクローン（λＢｂ２０４）と完全に一致することから，Ｃ－１遺伝子も脳で発現していると考えられる(1)．つまり，ボ

ンビキシン遺伝子は 機能しうるプロセスされた遺伝子と考えられる. 機能しうるプロセスされた遺伝子は，昆虫では最初の発見であり，昆虫以外の真核生物においてもその例は多くないが（2,47,63,64,67)，げっ歯類のインシュリンーＩ遺伝子は，プロセスされた遺伝子と考えられている(63).

５－５　　ボンビキシン遺伝子の進化

脊椎動物のインシュリン族遺伝子は，まず，セリンプロテアーゼ遺伝子から 祖先インシュリン遺伝子が分岐し進化した(14,66). 次に一つの祖先インシュリン遺伝子が重複し，それぞれインシュリン，ＩＧＦ－I，－II，リラキシンに進化したと考えられている. 事実，ヒトで，インシュリンとＩＧＦ－IIは11番染色体上に並んで存在し，ＩＧＦ－Iも，進化的に11番染色体と関連のある 12番染色体上に存在している(6,12,71).

免疫組織化学的手法により，カイコ脳内に抗ブタインシュリン抗体と反応する分子の存在が示されている (30,77). このカイコにおけるボンビキシン以外のインシュリン族分子の存在を考えあわせると，インシュリン族遺伝子としてのボンビキシン遺伝子は次のように進化したと考えられる. まず，ボンビキシン，インシュリン共通の祖先遺伝子が重複し，祖先インシュリン遺伝子と祖先ボンビキシン遺伝子に分かれる. 次に，祖先インシュリン遺伝子はそのままインシュリン族分子をコードする遺伝子（この遺伝子産物は抗インシュリン抗体と反応する）へと進化し，祖先ボンビキシン遺伝子は三つの各ファミリー遺伝子（これらの遺伝子産物は抗インシュリン抗体と反応しない）へ進化し，さらに遺伝子の重複を受けた. この際，ボンビキシン遺伝子では各ファミリー内の反復配列がよく保存されている（図１２）ことから，ボンビキシン遺伝子が各ファミリーに分化した後にプロセスされ，そして遺伝子の重複を受けた可能性，もしくは，ボンビキシン遺伝子がプロセスを受けゲノム中の異なる場所に挿入された後に各ファミリーに分化し，そして遺伝子の重複

を受けた可能性のいずれかが，その進化の過程として考えられる．

ボンビキシン遺伝子はプロセスされた遺伝子と考えられるが，現在までに単離した遺伝子の中から，親遺伝子に当たるイントロンをもつ遺伝子は見つかていない．この理由として，ただ親遺伝子がまだ単離されていない可能性のほか，ボンビキシン遺伝子の進化の過程で，次のようなことが起こった可能性も考えられる．親遺伝子がカイコゲノムから失われた，もしくは，プロセスされた遺伝子の進化速度が早いため，親遺伝子と塩基配列の相同性が低くなり，用いたプローブでは親遺伝子を単離できないといった可能性である．この単離できていない親遺伝子が，先の抗インシュリン抗体と反応する分子をコードする遺伝子であるかもしれない．

## ５－６　ボンビキシンの機能

エリサンと異なりカイコでは，幼虫4－5令の脱皮(69)，蛹－成虫(28)の変態に関する限り，ボンビキシンには前胸腺刺激活性がなく，現在カイコにおけるボンビキシンの生理機能は不明である．しかし，インシュリンとの相同性から，インシュリンの生理作用である血糖の調節にかかわっている可能性のほか，カイコＰＴＴＨがふ化の直前まで現われない(18)のに対し，ボンビキシンはカイコの発生卵中にごく初期から存在すること (24)，また，ボンビキシンが ＩＧＦ と 高いアミノ酸配列の相同性（Ａ鎖で 60-65%）を有する（表２）ことなどから，ボンビキシンはカイコの成長過程，とくに卵発生の過程において 重要な役割を果たしていると考えられる． 一方，インシュリンは，ショウジョウバエの成虫原器の成長や培養細胞の増殖を促すこと(19, 50)，また，スズメガの一種 Agrius cinglulata やオオカバマダラ Danaus plexippus の脂肪体よりトリグリセリド，ジグリセリド，脂肪酸の放出を促すこと (17) が知られている．これらのことから，インシュリン族ペプチドは，昆虫の発生過程で，代謝調節，細胞増殖や分化等の重要な機能を担っていることが予想される．

広く無脊椎動物(40,42,70)や原核生物(42)で，インシュリン族に属する分子の存在が，免疫学的もしくは生物学的方法により示されている．とくに，ショウジョウバエでは，ブタインシュリンと結合するインシュリンレセプターが存在し(22)，その遺伝子（ｃＤＮＡ）の一部がクローニングされ(54)，ヒトインシュリンレセプター遺伝子と塩基配列の相同性をもつことが示されている．このことは，ショウジョウバエには，インシュリンと構造的に似た分子の存在を示唆する．そして最近，軟体動物（リムネア）の成長に関与する脳中の神経分泌細胞において，インシュリン族ペプチド（ＭＩＰ）の存在がｃＤＮＡクローニングにより確かめられた(62)．これら一連の事実は，ボンビキシンをはじめとするインシュリン族ペプチドが脊椎動物，無脊椎動物，原核生物を問わず広く存在し生理的に重要な役割を担っていると考えられる．これらインシュリン族ペプチドの構造および機能を明らかにすることは，脳－消化管系ペプチドホルモンの機能や進化に貴重な知見を与え，新たな問題を提出するものである．

## 6 . 謝辞

本研究を遂行するにあたり，終始ご助言を賜った石崎宏矩教授，ならびに動物学第二講座の皆様方，大西英爾名誉教授，大沢省三教授ならびに動物学第一講座の皆様方，東京大学農学部鈴木昭憲教授，長沢寛道博士，オリゴヌクレオチドプローブを合成してくださった東京大学医科学研究所山本正幸助教授，竹田忠行博士，カイコの飼育に当たられた相馬甲子男元技官，公保郁子技官に感謝します．

# 7 . 引用文献


1. Adachi,T., Takiya, S., Suzuki, Y., Iwami, M., Takahashi, S. Y., Ishizaki, H., Nagasawa, H. and Suzuki, A. 1989. cDNA structure and expression of bombyxin, an insulin-like brain secretory peptide of the silkmoth Bombyx mori. J. Biol. Chem. 264, 7681-7685.
2. Andersen, R. D., Birren, B. W., Taplitz, S. J. and Herschman, H. R., 1986. Rat metallothionein-1 structural gene and three pseudogenes, one of which contains 5'-regulatory sequences. Mol. Cell. Biol., 6, 302-314.
3. Bell, G. I., Swain, W. F., Pictet, R., Cordell, B., Goodman, H. M. and Rutter, W. J., 1979. Nucleotide sequence of a cDNA clone encoding human preproinsulin. Nature, 282, 525-527.
4. Bell, G. I., Pictet, R., Rutter, W., Cordell, B., Tischer, E. and Goodman, H. M., 1980. Sequence of the human insulin gene. Nature, 284, 26-32.
5. Bell, G. I., Merryweather, J. P., Sanchez-Pescador, R., Stempien, M. M., Priestley, L., Scott, J. and Rall, L. B., 1984. Sequence of a cDNA clone encoding human pre-proinsulin-like growth factor II. Nature, 310, 775-777.
6. Bell, G. I., Gerhard, D. G., Fong, N. M., Sanchez-Pescador, R. and Rall, L. B., 1985. Isolation of the human insulin-like growth factor genes: insulin-like growth factor II and insulin gene are contiguous. Proc. Natl. Acad. Sci. U. S. A., 82, 6450-6454.

7. Blundell, T. L., Dodson, G. G., Dodson, E., Hodgkin, D. C. and Vijayan., 1970. X-ray analysis and the structure of insulin. Rec. Progr. Horm. Res., 27, 1-40.

8. Blundell, T. L. and Wood, S. P., 1975. Is the evolution of insulin Darwinian or due to selectively neutral mutation? Nature, 257, 197-203.

9. Blundell, T. L. and Humbel, R. E., 1980. Hormone families: pancreatic hormones and homologous growth factors. Nature, 287, 781-787.

10. Blundell, T. and Wood, S., 1982. The conformation, flexibility, and dynamics of polypeptide hormones. Rec. Progr. Horm. Res., 51, 123-154.

11. Bollenbacher, W. E. and Granger, N. A., 1985. Endocrinology of the prothoracicotropic hormone. In Comprehensive insect physiology, biochemistry and pharmacology. (Eds. G. A. Kerkut and L. I. Gilbert), Pergamon Press, Oxford.

12. Brissenden, J. E., Ullrich, A. and Francke, U., 1984. Human chromosomal mapping of genes for insulin-like growth factors I and II and epidermal growth factor. Nature, 310, 781-784.

13. Cavener, D. R., 1987. Comparison of the consensus sequence flanking translational start sites in Drosophila and vertebrates. Nucl. Acids Res., 15, 1353-1361.

14. Chan, S. J, Kwok, S. C. M. and Steiner, D. F., 1981. The biosynthesis of insulin: some genetic and evolutionary aspects. Diabetes Care, 4, 4-10.

15. Chan, S. J., Emdin, S. O., Kwok, S. C. M., Kramer, J. M., Falkmer, S. and Steiner, D. F., 1981. Messenger RNA sequence and primary structure of preproinsulin in a primitive vertebrate, the atlantic hagfish. J. Biol. Chem., 256, 7595-7602.

16. Chan, S. J., Episkopou, V., Zeitlin, S., Karathanasis, S. K., MacKrell, A., Steiner, D. F. and Efstratiadis, A., 1984. Guinea pig preproinsulin gene: an evolutionary compromise? Proc. Natl. Acad. Sci. U.S.A., 81, 5046-5050.

17. Chang, F., 1974. Effects of vertebrate adipokinetic hormones on the rate of in vitro lipid release in insects. Comp. Biochem. Physiol., 49B, 567-578.

18. Chen, J. H., Fugo, H., Nakajima, M., Nagasawa, H. and Suzuki, A., 1987. Neurohormones in developing embryos of the silkworm, Bombyx mori: the presence and characteristics of prothoracicotropic hormone B. J. Insect Physiol., 33, 407-411.

19. Davis, K. T. and Shearn, A., 1977. In vitro growth of imaginal disks from Drosophila melanogaster. Science, 196, 438-439.

20. Dull, T. J., Gray, A., Hayflick J. S. and Ullrich, A., 1984. Insulin-like growth factor II precursor gene organization in relation to insulin gene family. Nature, 310, 777-781.

21. Feinberg, A. P. and Vogelstein, B., 1983. A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. Anal. Biochem., 132, 6-13.

22. Fernandez-Almonacid, R. and Rosen, O. M., 1987. Structure and ligand specificity of the Drosophila melanogaster insulin receptor. Mol. Cell. Biol., 7, 2718-2727.

23. Frischauf, A-M., Lehrach, H., Poustka, A. and Murray, N., 1983. Lambda replacement vectors carrying polylinker sequences. J. Mol. Biol., 170, 827-842.

24. Fugo, H., Hua, J., Nakajima, M., Nagasawa, N. and Suzuki, A., 1987. Neurohormones in developing embryos of the silkworm, Bombyx mori: the presence and characteristics of prothoracicotropic hormone-S. J. Insect Physiol., 33, 243-248.

25. Henikoff, S., 1984. Unidirectional digestion with exonuclease III creates targeted breakpoints for DNA sequencing. Gene, 28, 351-359.

26. Hudson, P., Haley, J., John, M., Cronk, M., Crawford, R., Haralambidis, J., Tregear, G., Shine, J. and Niall, H., 1983. Structure of a genomic clone encoding biologically active human relaxin. Nature, 301, 628-631.

27. Hultmark, D., Klemenz, R. and Gehring, W. J., 1986. Translational and transcriptional control elements in the untranslated leader of the heat-shock gene hsp22. Cell, 44, 429-438. (1986).

28. Ishizaki, H., Mizoguchi, A., Fujishita, M., O'oka, H., Kataoka, H., Isogai, A., Nagasawa, H., Tamura, S. and Suzuki, A., 1983. Species specificity of the insect prothoracicotropic hormone (PTTH): the presence of Bombyx- and Samia-specific PTTHs in the brain of Bombyx mori. Develop. Growth and Differ., 25, 593-600.

29. Ishizaki, H. and Suzuki, A., 1984. Prothoracicotropic hormone of Bombyx mori. In Biosynthesis, Metabolism and Mode of Action of Invertebrate Hormones (Eds. J. Hoffmann and M. Porchet), pp.63-77. Springer-Verlag, Heidelberg.

30. Ishizaki, H., Mizoguchi, A., Hatta, M., Suzuki, A., Nagasawa, H., Kataoka, H., Isogai, A., Tamura, S., Fujino, M. and Kitada, C., 1987. Prothoracicotropic hormone (PTTH) of the silkmoth, Bombyx mori: 4K-PTTH. In Molecular Entomology (Ed. J. H. Law), pp.119-128. Alan R. Liss, New York.

31. Ishizaki, H. and Suzuki, A. 1988. An insect brain peptides as a member of insulin family. Horm. Metabol. Res. 20, 426-429.

32. Iwami, M., Kawakami, A., Ishizaki, H., Takahashi, S. Y., Adachi, T., Suzuki, Y., Nagasawa, H. and Suzuki, A. 1989. Cloning of a gene encoding bombyxin,an insulin-like brain secretory peptide of the silkmoth Bombyx mori with prothoracicotropic activity. Devel. Growth and Differ., 31, 31-37.

33. Iwami, M. 1990. The gene encoding bombyxin, a brain secretory peptide of Bombyx mori: structure and expression. In Molting and Metamorphosis (Eds. E. Ohnishi and H. Ishizaki), in press. Japan Scientific Societies Press, Tokyo, and Springer-Verlag, Heidelberg.

34. Jansen, M., van Schaik, F. M. A., Ricker, A. T., Bullock, B., Woods, D. E., Gabbay, K. H., Nussbaum, A. L., Sussenbach, J. S. and Van den Brande, J. L., 1983. Sequence of cDNA encoding human insulin-like growth factor I precursor. Nature, 306, 609-611.

35. Jhoti, H., McLeod, A. N., Blundell, T. L., Ishizaki, H., Nagasawa, H. and Suzuki, A., 1987. Prothoracicotropic hormone has an insulin-like tertialy structure. FEBS Letters, 219, 419-425.

36. Kafatos, F. C., Mitsialis, S. A., Nguyen, H. T., Spoerel, N., Tsitilou, S. G. and Mazur, G. D., 1987. Evolution of the structural genes and regulatory elements for the insect chorion. In Development as an Evolutionary Process. (Eds. R. A. Rudolf and R. C. Elizabeth), pp.161-178. Alan R. Liss, New York.

37. Kahn, C. R., Baird, K. L., Flier, J. S., Grunfeld, C., Harmon, J. T., Harrison, L. C., Karlsson, F. A., Kasuga, M., King, G. L., Lang, U. C., Podskalny, J. M. and Obberghen, E. V., 1981. Insulin receptors, receptor antibodies, and the mechanism of insulin action. Rec. Progr. Horm. Res., 37, 477-538.

38. Kawakami, A., Iwami, M., Nagasawa, H., Suzuki, A. and Ishizaki, H. 1989. Structure and organization of four clustered genes that encode bombyxin, an insulin-related brain secretory peptide of the silkmoth Bombyx mori. Proc. Natl. Acad. Sci. U.S.A. 86,

39. Kimura, M, 1981. Estimation of evolutionary distances between homologous nucleotide sequences. Proc. Natl. Acad. Sci. U.S.A., 78, 454-458

40. Kramer, K. J., 1985. Vertebrate hormones in insects. In Comprehensive insect physiology, biochemistry, and pharmacology Vol.7. (Eds. G. A. Kerkut and L. I. Gilbert), pp.511-536. Pergamon Press, Oxford.

41. Lathe, R., 1985. Synthetic oligonucleotide probes deduced from amino acid sequence data: Theoretical and practical considerations. J. Mol. Biol., 183, 1-12.

42. LeRoith, D., Delahunty, G., Wilson, G. L., Roberts, C. T. Jr., Shemer, J., Hart, C., Lesniak, M. A., Shiloach, J. and Roth, J., 1986. Evolutionary aspects of the endocrine and nervous systems. Rec. Progr. Horm. Res., 42, 549-587.

43. Maruyama, T., Gojobori, T., Aota, S. and Ikemura, T., 1986. Codon usage tabulated from the GenBank genetic sequence data. Nucl. Acids Res., 14, r151-r197.

44. Maruyama, K., Hietter, H., Nagasawa, H., Isogai, A., Tamura, S., Suzuki, A. and Ishizaki, H., 1988. Isolation and primary structure of bombyxin-IV, a novel molecular species of bombyxin from the silkworm, Bombyx mori.

45. Mason, P. J. and Williams, J. G., 1985. Hybridisation in the analysis of recombinant DNA. In Nucleic acid hybridization: a practical approach. (Eds. B. D. Hames and S. J. Higging), IRL Press, Oxford.

46. Meyts, P. D., Van Obberghen, E., Roth, J., Wollmer, A. and Brandenburg, D., 1978. Mapping of the residues responsible for the negative cooperativity of the receptor-binding region of insulin. Nature, 273, 504-509.

47. McCarrey, J. R. and Thomas, K., 1987. Human testis-specific PGK genes lacks introns and possesses characteristics of a processed gene. Nature, 326, 501-505.

48. Mizoguchi, A., Ishizaki, H., Nagasawa, H., Kataoka, H., Isogai, A., Tamura, S., Suzuki, A., Fujino, M. and Kitada, C., 1987. A monoclonal antibody against a synthetic fragment of bombyxin (4K-prothoracicotropic hormone) from the silkmoth, Bombyx mori: characterization and immunohistochemistry. Mol. Cell. Endocr., 51, 227-235.

49. Mizusawa, S., Nishimura, S. and Seela, F., 1986. Improvement of the dideoxy chain termination method of DNA sequencing by use of deoxy-7-deazaguanosine triphosphate in place of dGTP. Nucl. Acids Res., 14, 1319-1324.

50. Mosna, G., 1981. Insulin can completely replace serum in Drosophila melanogaster cell cultures in vitro. Experientia, 37, 466-467.

51. Nagasawa, H., Kataoka, H., Isogai, A., Tamura, S., Suzuki, A. Ishizaki, H., Mizoguchi, A., Fujiwara, Y. and Suzuki, A., 1984. Amino-terminal amino acid sequence of the silkworm prothoracicotropic hormone: homology with insulin. Science, 226, 1344-1345.

52. Nagasawa, H., Kataoka, H., Isogai, A., Tamura, S., Suzuki, A., Mizoguchi, A., Fujiwara, Y., Suzuki, A., Takahashi, S. Y. and Ishizaki, H., 1986. Amino acid sequence of a prothoracicotropic hormone of the silkworm Bombyx mori. Proc. Natl. Acad. Sci. U. S. A., 83, 5840-5843.

53. Nagasawa, H., Maruyama, K., Sato, B., Hietter, H., Kataoka, H., Isogai, A., Tamura, S., Ishizaki, H., Senba, T. and Suzuki, A., 1988. Structure and synthesis of bombyxin from the silkworm, Bombyx mori. In Peptide Chemistry 1987. (Eds. T. Shiba and S. Sakakibara), pp.123-126. Protein Research Foundation, Osaka.

54. Nishida, Y., Hata, M., Nishizuka, Y., Rutter, W. and Ebina, Y., 1986. Cloning of a Drosophila cDNA encoding a polypeptide similar to the human insulin receptor precursor. Biochem. Biophys. Res. Com., 141, 474-481.

55. Pullen, R. A., Lindsay, D. G., Wood, S. P., Tickle, I. J., Blundell, T. L., Wollmer, A., Krail, G., Brandenburg, D., Zahn, H., Gliemann, J. and Grammeltoft., 1976. Receptor-binding region of insulin. Nature, 259, 369-373.

56. Reed, K. C. and Mann, D. A., 1985. Rapid transfer of DNA from agarose gels to nylon membranes. Nucl. Acids Res., 13, 7207-7221.

57. Renan, M. J., 1987. Conserved 12-bp element downstream from mRNA polyadenylation sites. Gene, 60, 245-254.

58. Rogers, J. 1983. Retroposons defined. Nature, 301, 460.

59. Sakai, N., Mori, S., Izumi, S., Haino-Fukushima, K., Ogura, T., Maekawa, H. and Tomino, S., 1988. Structures and expression of mRNAs coding for major plasma proteins of Bombyx mori. Biochem. Biophys. Acta, 949, 224-232.

60. Sanger, F., Nicklen, S. and Coulson, A. R., 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. U. S. A., 74, 5463-5467.

61. Sharp, P. A., 1983. Conversion of RNA to DNA in mammals: Alu-like elements and pseudogenes. Nature, 301, 471-472.

62. Smit, A. B., Vreugdenhil, R., Ebberink, R. H. M., Geraerts, W. P. M., Klootwijk, J. and Joosse, J., 1988. Growth-controlling molluscan neurons produce the precursor of an insulin-related peptide. Nature, 331, 535-538.

63. Soares, M. B., Schon, E., Henderson, A., Karathanasis, S. K., Cate, R., Zeitlin, S., Chirgwin, J. and Efstratiadis, A. 1985. RNA-mediated gene duopication: the rat preproinsulin I gene is a functional retroposon. Mol. Cell. Biol. 5, 2090-2103.

64. Stein, J. P., Munjaal, R. P., Lagace, L., Lai, E. C., O'Malley, B. W. and Means, A. R., 1983. Tissue-specific expression of a chicken calmodulin pseudogene lacking intervening sequences. Proc. Natl. Acad. Sci. U. S. A., 80, 6485-6489.

65. Steiner, D. F., 1978. On the role of the proinsulin C-peptide. Diabetes, 27 (Suppl. 1), 145-148.

66. Steiner, D. F., Chan, S. J, Welsh, J. M. and Kwok, S. C. M., 1985. Structure and evolution of the insulin gene. Ann. Rev. Genet., 19, 463-484.

67. Stutz, F. and Spohr, G., 1987. A processed gene coding for a sarcomeric actin in Xenopus laevis and Xenopus tropicalis. EMBO J., 6, 1989-1995.

68. Suzuki, A., Nagasawa, H. and Ishizaki, H., 1989. Isolation and structure of peptide hormones from the silkworm Bombyx mori: bombyxins-peptides structurally related to insulin. Pure Appl. Chem., 61, 349-352.

69. Suzuki, C. and Ishizaki, H., 1986. Prothoracicotropic hormone bioassay: Bombyx larva assay. Int. J. Invert. Reprod. Devel., 10, 256-274.

70. Thorpe, A. and Duve, H., 1984. Insulin- and glucagon-like peptides in insects and molluscs. Mol. Physiol., 5, 235-260.

71. Tricoli, J. V., Rall, L. B., Scott, J., Bell, G. I. and Shows, T. B., 1984. Localization of insulin-like growth factor genes to human chromosomes 11 and 12. Nature, 310, 784-786.

72. Turner, A. J., 1986. Processing and metabolism of neuropeptides. In Essays in Biochemistry 22. (Eds. R. D. Marchall and K. F. Tipton), pp. 69-119. Academic Press, London.

73. Vanin, E. F., 1985. Processed pseudogenes: characteristics and evolution. Ann. Rev. Genet., 19, 253-272.

74. Von Heijne, G., 1983. Patterns of amino acids near signal-sequence cleavage sites. Eur. J. Biochem., 133, 17-21.

75. Watson, M. E. E., 1984. Compilation of published signal sequences. Nucl. Acids Res., 12, 5145-5164.

76. Wood, S. P., Blundell, T. L., Wollmer, A., Lazarus, N. R. and Neville, W. J., 1975. The relation of conformation and association of insulin to receptor binding; X-ray and circular-dichroism studies on bovine and hystricomorph insulins. Eur. J. Biochem., 55, 531-542.

77. Yanisch-Perron, C., Vieira, J. and Messing, J., 1985. Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. Gene, 33, 103-119.

78. Yui, R., Fujita, T. and Ito, S., 1980. Insulin-, gastrin-, pancreatic polypeptide-like immunoreactive neurons in the brain of the silkworm, Bombyx mori. Biomed. Res., 1, 42-46.

表1： ヌクレオチド座位当りの進化的距離（K）

| Comparison | Chain or Peptide | | |
|---|---|---|---|
| | A | B | C |
| bombyxin A-1 vs. B-1 | 0.47 | 0.39 | 0.64 |
| A-1 vs. C-1 | 0.40 | 0.75 | 0.54 |
| B-1 vs. C-1 | 0.40 | 1.02 | 0.56 |
| human insulin vs. rat insulin-I | 0.07 | 0.20 | 0.44 |

表２：　インシュリン族ペプチドＡ鎖間のアミノ酸配列の相同性（％）

| | | Bombyxin | | | Insulin | IGF-I |
|---|---|---|---|---|---|---|
| | | A-1 | B-1 | C-1 | | |
| | A-1 | | | | | |
| Bombyxin | B-1 | 80 | | | | |
| | C-1 | 80 | 80 | | | |
| Insulin | | 50 | 45 | 45 | | |
| IGF-I | | 60 | 65 | 60 | 55 | |
| IGF-II | | 60 | 65 | 55 | 60 | 80 |

# ８．図の説明

図１

オリゴヌクレオチドプローブ

ボンビキシン遺伝子を単離するために合成した二種類のオリゴヌクレオチドプローブ．上段はプローブＡを，下段はプローブＢを示している．各プローブの上行は，ボンビキシン Ａ鎖（A-chain）またはＢ鎖（B-chain）のアミノ酸配列を 一文字表記で，次の行（ｍＲＮＡ）は，予想されるｍＲＮＡの塩基配列を，続いて，合成したプローブの塩基配列を示している．

図２

プローブＡに陽性のクローンのサザンハイブリダイゼーション

プローブＡによるスクリーニングにより得られた６個のクローンをSalIで切断しアガロース電気泳動の後，メンブレンフィルターに転写し，プローブＡ（ａ），または，プローブＢ（ｂ）とサザンハイブリダイゼーションを行った．レーン１，λ４Ｋ１０１；レーン２，λ４Ｋ１０５；レーン３，λ４Ｋ１０７；レーン４，λ４Ｋ１０８；レーン５，λ４Ｋ１１１；レーン６，λ４Ｋ１１２．レーン６の右に分子量マーカー（HindIII切断λＤＮＡ）の位置を示した．プローブとのハイブリダイゼーションの条件はスクリーニングの条件に同じ．

図３

ボンビキシン遺伝子の塩基配列決定の戦略

上段はボンビキシンＡ－１［Bombyxin A-1（λ４Ｋ１１２）］遺伝子の，下段はボンビキシンＣ－１［Bombyxin C-1（λ４Ｋ４０３）］遺伝子の塩基配列決定の戦略を示している．Ｈ，Ｐ，Ｋ，Ｅは順に HindIII，PstI，KpnI，EcoRI による切断部位を，矢印の方向と

長さは塩基配列を決定した向きと長さを示している．また，遺伝子地図中の CAT，TATA，AATAAA，S，B，C，Aは順に，CAT－box，TATA-box，ポリ(A)付加シグナル，シグナルペプチド，B鎖，Cペプチド，A鎖を表わしている．

図４

ボンビキシンA－１遺伝子の塩基配列

ボンビキシンA－１遺伝子 HindIII断片1045塩基対の 塩基配列(32)．アミノ酸配列の順番は，B鎖N端を１とした．シグナルペプチド，B鎖，Cペプチド，A鎖をそれぞれ括弧で囲んだ．星印は翻訳停止コドンを，陰を付けた配列は，CAT-box (GGCAATGT)，TATA-box (TATATAA)を，細い下線を引いた配列は，ポリ(A)付加 シグナル (AATAAA)を，波線を引いた配列は12塩基対の保存配列 (ATACAGTTTTT)を，また，箱で囲んだ配列 (TTCTTC) は６塩基対の反復配列を示している．太い下線を引いた配列 (CACATTC，27参照) は，primer extension法により決定された(38)転写開始部位をしめす．

図５

予想されるボンビキシンの生合成過程

黒三角はトリプシン様エンドペプチダーゼにより，切断される部位を示している．

図６

ボンビキシンA－１遺伝子の発現とゲノムあたりのコピー数

a）プローブaを用いたノーザンハイブリダイゼーション．レーン１，絹糸腺由来 ポリ(A)RNA (3.3μg)；レーン２，幼虫脳由来 ポリ(A)RNA(5μg)．b）プローブbを用いた ノーザンハイブリダイゼーション．レーン１，２は，a）に同じ．c）カイコゲノムDNA(5μg)をBamHI (レーン１)，もしくは，HindIII (レーン２) で切断し，プローブaとハイブリダイズさせたサザンハイブ

リダイゼーション．レーンの左側に，分子量マーカー（HindIII 切断λＤＮＡ），右側に，主要なHindIII断片の大きさをキロ塩基対（ｋｂ）で示している．ｄ）サザンハイブリダイゼーションおよびノーザンハイブリダイゼーションに用いたプローブ．Pre は，シグナルペプチドを示している．他は図３に同じ．

図７

ボンビキシンＣ－１遺伝子の塩基配列

ボンビキシンＣ－１遺伝子HindIII/KpnI断片 456塩基対の塩基配列．アミノ酸配列の順番は，Ｂ鎖Ｎ端を１とした．シグナルペプチド，Ｂ鎖，Ｃペプチド，Ａ鎖をそれぞれ括弧で囲んだ．星印は停止コドンを，下線を引いた配列は，ポリ(Ａ)付加 シグナル (AATAAA) を，箱で囲んだ配列 (5'側，AATATT；3'側，AATATAT) は 反復配列を，また，太い下線を引いた配列(CACATCC)は 予想される転写開始位置(27,38)を示している．

図８

ボンビキシン遺伝子間のハープロット解析

（ａ）ボンビキシンＡ－１とＢ－１(38)間，（ｂ）Ｃ－１とＡ－１間，（ｃ）Ｃ－１とＢ－１間のハープロット解析．Ｓ，Ｂ，Ｃ，Ａは順に，シグナルペプチド，Ｂ鎖，Ｃペプチド，Ａ鎖を示している．ハープロット解析に用いたパラメーターは，チエックサイズを15，一致させる塩基数を11に設定した．

図９

インシュリン族ペプチドのアミノ酸配列の比較

ボンビキシンＡ－１，Ａ－２ (38)，Ａ－３ (38)，Ｂ－１ (38)，Ｂ－２(38)，Ｃ－１，ボンビキシン－I (51)，－II (52, 53)，－III (51)，－IV (35,44)，－V (35)，ヒトインシュリン (4)，モルモットインシュリン (16)，ヒトＩＧＦ－I (34)，－II(5)，および

ヒトリラキシン(26)のアミノ酸配列の比較．ヒトインシュリンの各ドメインのN端を１とし，以下順次アミノ酸残基に番号をつけた．ギャップは対応するアミノ酸残基が存在しないことを，ダッシュは上のアミノ酸残基と同じ配列であることを示している．ボンビキシンを含むインシュリン族ペプチドすべてに共通なアミノ酸残基は桃色で，ボンビキシンペプチドに共通なアミノ酸残基は青色で，ボンビキシンを除くインシュリンペプチドに共通なアミノ酸残基は黄色で示した．

図１０

ボンビキシンとインシュリン両遺伝子間のハープロット解析

（ａ）ボンビキシンＡ－１とヒトインシュリン遺伝子(3)間，（ｂ）ボンビキシンＢ－１(38)とヒトインシュリン遺伝子間，（ｃ）ボンビキシンＣ－１とヒトインシュリン遺伝子間のハープロット解析．縦軸にヒトインシュリン遺伝子，横軸にボンビキシン遺伝子の塩基配列をとっている．Ｓ，Ｂ，Ｃ，Ａは順に，シグナルペプチド，Ｂ鎖，Ｃペプチド，Ａ鎖を示している．ハープロット解析に用いたパラメーターは，図８と同じである．

図１１

インシュリン族遺伝子の遺伝子構成

ボンビキシンＡ－１，ヒトインシュリン(5)，ヒトＩＧＦ－I(7)，－II(21)遺伝子の遺伝子構成(32)．影をつけた部分は，生合成の過程で除去されるペプチドを示している．各ペプチド部分の数字は各ペプチドをコードする塩基数を，ATG，TAG，TGA は停止コドンを，黒三角はイントロンを示している．KR，RR，Rは トリプシン様エンドペプチダーゼにより切断されるアミノ酸残基を一文字表記により示したものである．

図１２

ボンビキシン遺伝子に存在する反復配列

ボンビキシンＡ－１，Ａ－２，Ａ－３，Ｂ－１，Ｂ－２，Ｃ－１遺伝子の予想される転写領域の両側に存在する反復配列．反復配列は箱にいれて示した．また，Ａファミリー遺伝子中に存在するポリ(Ａ)配列の残存配列に下線を引いた．

図 1

Probe A (51mer, A-chain Probe)

```
     4                                                    20
H - D  E  C  C  L  R  P  C  S  V  D  V  L  L  S  Y  C   -OH   Bombyxin-II A-chain

5'- GACGAGUGCUGCCUCAGACCCUGCAGCGUCGACGUCCUCCUCAGCUACUGC -3'   mRNA

3'- CTGCTCACGACGGAGTCTGGGACGTCGCAGCTGCAGGAGGAGTCGATGACG -5'   Probe A
```

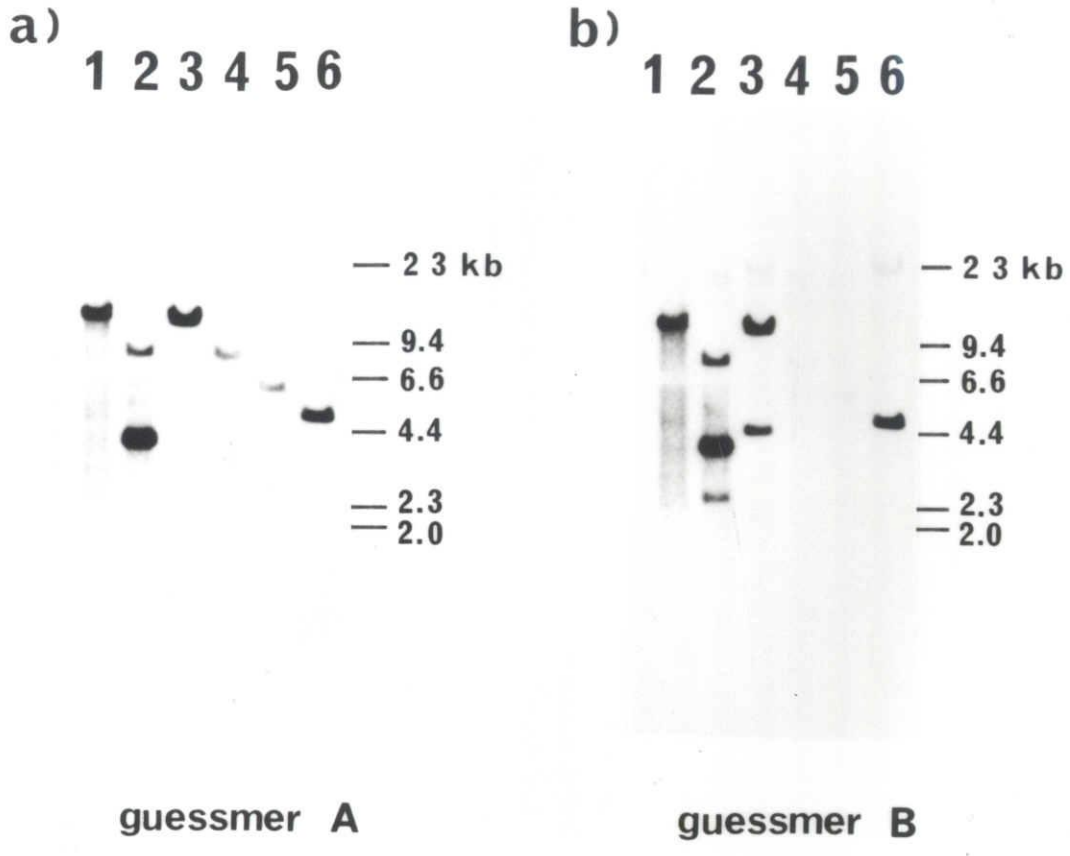
Probe B (51mer, 16 mixture, B-chain Probe)

```
     8                                                    24
H - T  Y  C  G  R  H  L  A  R  T  L  A  D  L  C  W  E   -OH   Bombyxin-II B-chain

5'- ACCUACUGCGGACGCCACCUGGCCCGCACCCUGGCCGACCUGUGCUGGGAG -3'   mRNA
              C           U           U  U

3'- TGGATGACGCCTGCGGTGGACCGGGCGTGGGACCGGCTGGACACGACCCTC -5'   Probe B
              G           A           A  A
```

図2


a)
1 2 3 4 5 6
— 23 kb
— 9.4
— 6.6
— 4.4
— 2.3
— 2.0
guessmer A
b)
1 2 3 4 5 6
— 23 kb
— 9.4
— 6.6
— 4.4
— 2.3
— 2.0
guessmer B

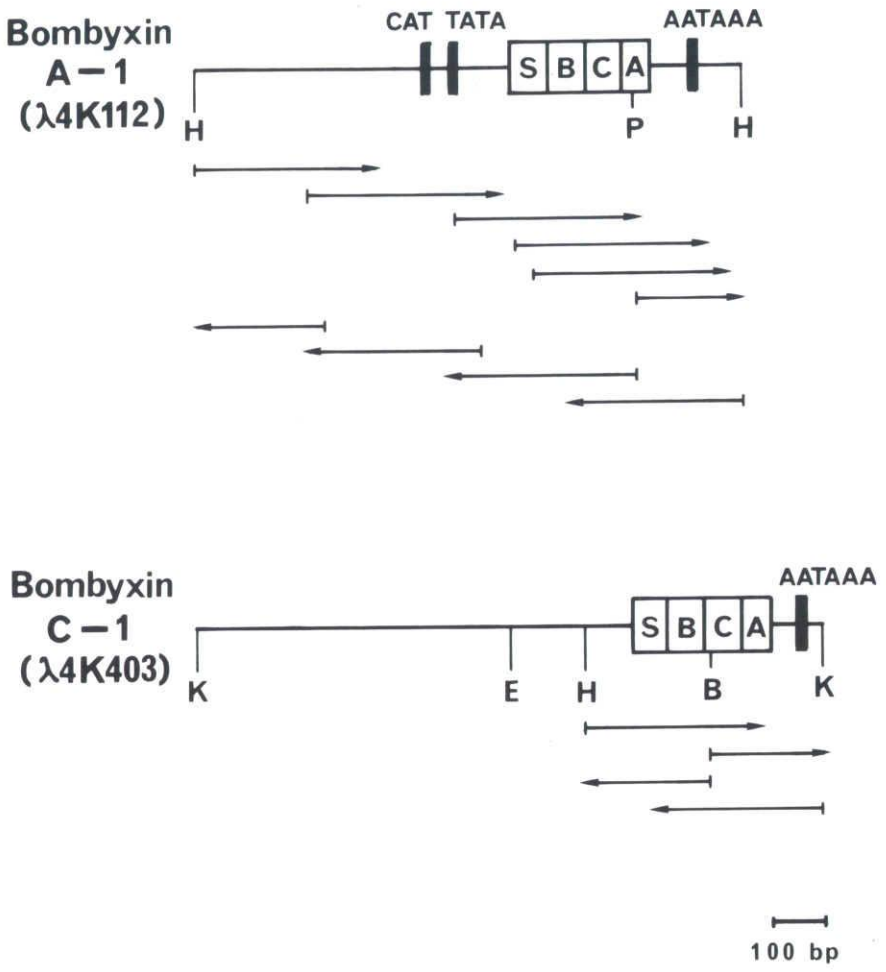
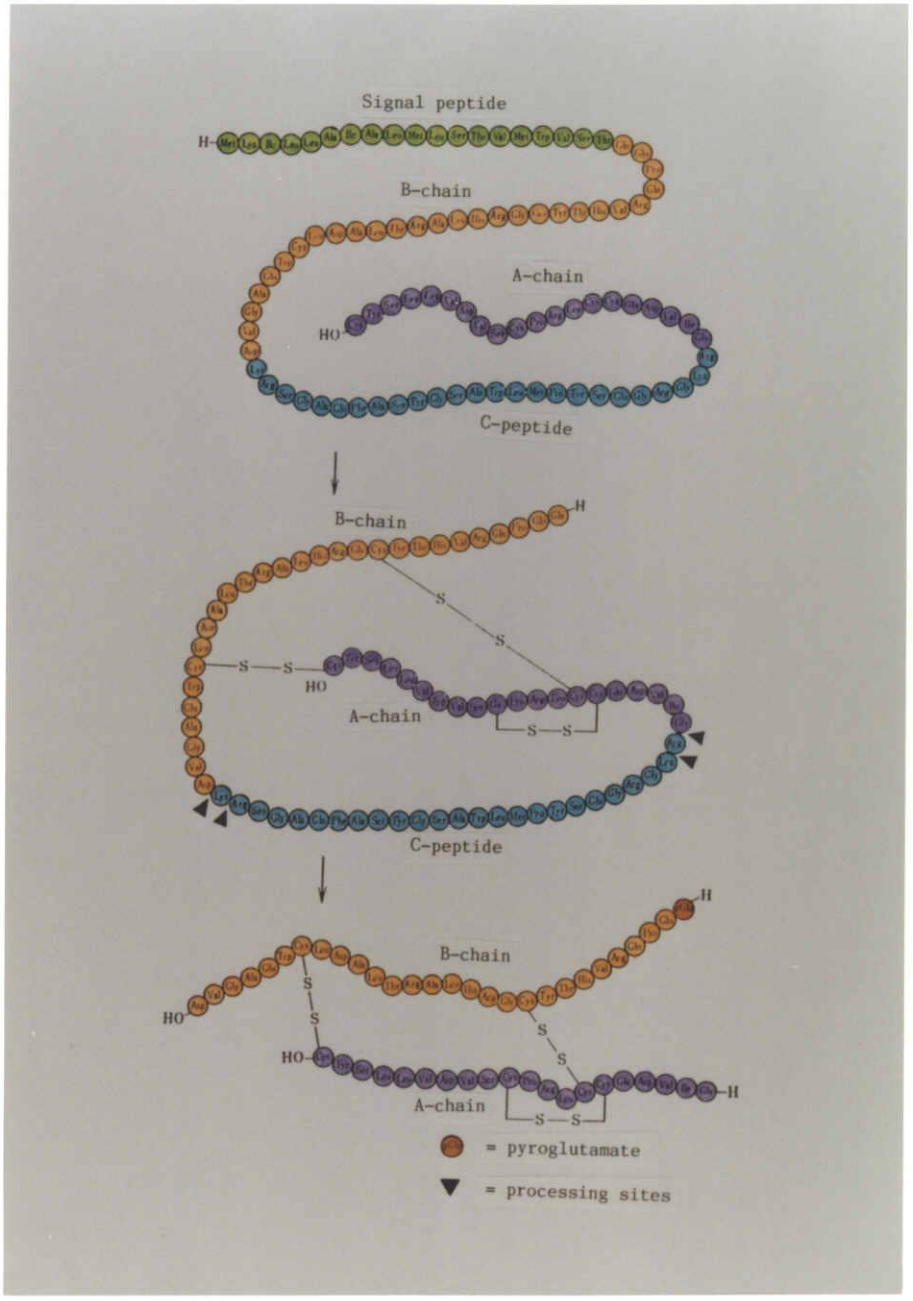

Bombyxin
A−1
(λ4K112)
CAT TATA
AATAAA
S B C A
H
P
H
Bombyxin
C−1
(λ4K403)
AATAAA
S B C A
K
E
H
B
K
100 bp

図4


5'-AAGCTTGTGGAACCTCGCAACAAACGAGACCGTCTTAACCTAAGGCTTCAAAGAATGAT 59
TTCAACGTTGTAACCCGGTCACCACTTGAAACAAGCAGTTTCTCTATGGTAAAATTTTCC 119
ACTACCAATCTTTTGAAATGTTTCAAGCTATCAAAGCAATTGGACCACGCCATACTTTCA 179
CAAATAAATCATAAAACATGATTACTACAGCTCTGATGAAGTCATAAACGCTCGCTTCCA 239
GCCACGCGTGCATCGAGCGTTATGAACTGGTTCTAAGTTCACCACTGATACCGGATGATG 299
CTTACTTTGTCTCCTAGTTACTACCTACTACATAATGAAGATGACAAATATTTATAAATT 359
TTGCTAGGCGTTCAAAGTTAAAAGTTGCGTCAAGTAAATGTACAACATTTAATTCAGTTC 419
ATTGGTTAATTATTGGAACCTAGGCAATGTCATATAAATCACATGCCCATGTGTTATTAT 479
ATAAATATTCATTAACTAATCGTCCATATATAAAGGAAGGGTTGCTTAGCTTTCTTCACA 539
***
TTCACGTTTCATCTGTGGCATCAGCGAGACGTCCTCCGAAACACACTTTAGTCGCCAAAC 599
SIGNAL PEPTIDE
-19 -10 1
MetLysIleLeuLeuAlaIleAlaLeuMetLeuSerThrValMetTrpValSerThrGln
ATGAAGATACTCCTTGCTATTGCATTAATGTTGTCAACAGTAATGTGGGTGTCAACACAA 659
B-CHAIN
10 20
GlnProGlnArgValHisThrTyrCysGlyArgHisLeuAlaArgThrLeuAlaAspLeu
CAGCCACAAAGAGTGCACACGTACTGCGGGCGTCACTTGGCTCGCACTCTGGCCGACCTG 719
C-PEPTIDE
30 40
CysTrpGluAlaGlyValAspLysArgSerGlyAlaGlnPheAlaSerTyrGlySerAla
TGCTGGGAAGCGGGCGTGGACAAGCGCAGCGGTGCTCAGTTCGCGAGCTACGGCTCCGCG 779
A-CHAIN
50 60
TrpLeuMetProTyrSerGluGlyArgGlyLysArgGlyIleValAspGluCysCysLeu
TGGCTGATGCCGTACTCGGAAGGACGCGGCAAACGAGGCATCGTGGATGAGTGCTGTCTC 839
70
ArgProCysSerValAspValLeuLeuSerTyrCys***
AGACCCTGCAGCGTGGACGTGCTTCTGTCGTACTGTTAGACCATTCCTTTACCGAATGGT 899
TCGTTCTCAATACATTTGGAAATAAAAAACCGTAAATAATAATAAAAAGTTCTTCTTTTC 959
CTGAACCACAATTGAAAGTGTTGACTGGAGATTTAGATACAGTTTTTTAGACATCATTTA 1019
GATGATTTTGATTGGCAAGCAAGCTT-3' 1045

図5


Signal peptide
B-chain
A-chain
C-peptide
B-chain
A-chain
C-peptide
B-chain
A-chain
= pyroglutamate
= processing sites

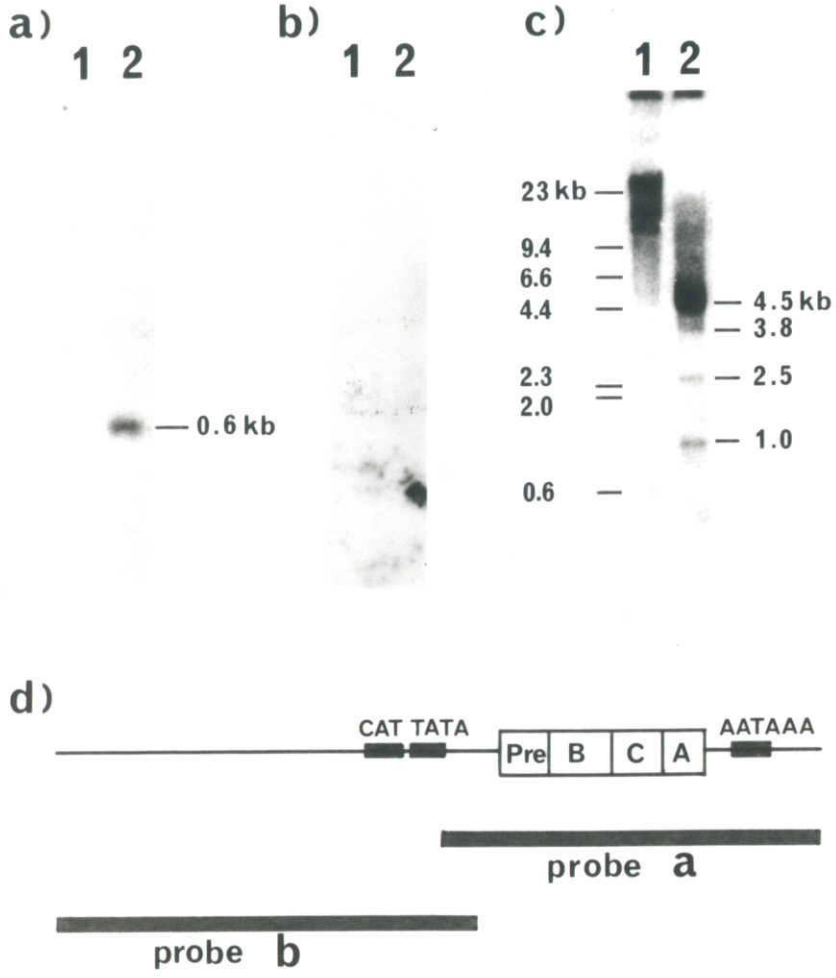

a)
1 2
0.6 kb
b)
1 2
c)
1 2
23 kb
9.4
6.6
4.4
2.3
2.0
0.6
4.5 kb
3.8
2.5
1.0
d)
CAT TATA
Pre
B
C
A
AATAAA
probe a
probe b

図 7


5'-AAGCTTCCATATAAAATGAGAGACTTTCGCACGCTCGCAATATTCA 46
-19
*** MetLysLeu -17
CATCCTGCCATCACAACGAGTAGCACGTGAGGAATCATCATGAAACTG 94
SIGNAL PEPTIDE
-10 -1
ValMetLeuLeuValValValSerAlaMetLeuValLeuGlyGlyAla -1
GTCATGCTCCTTGTTGTCGTTTCTGCCATGCTCGTGTTAGGAGGAGCA 142
B-CHAIN
1 10
GlnThrAlaSerGlnPheTyrCysGlyAspPheLeuAlaArgThrMet 16
CAAACAGCCAGCCAATTTTATTGTGGAGACTTTTTGGCGCGTACAATG 190
20 30
SerSerLeuCysTrpSerAspMetGlnLysArgSerGlySerGlnTyr 32
TCCAGCCTGTGTTGGTCGGACATGCAGAAACGAAGCGGATCCCAGTAC 238
C PEPTIDE
40
AlaGlyTyrGlyTrpProTrpLeuProProPheSerSerSerArgGly 48
GCGGGCTACGGCTGGCCGTGGCTGCCCCCCTTCTCCTCGTCTCGTGGG 286
A-CHAIN
50 60
LysArgGlyIleValAspGluCysCysTyrArgProCysThrIleAsp 64
AAACGCGGTATTGTCGATGAGTGCTGCTACAGACCCTGCACAATAGAC 334
70
ValLeuMetSerTyrCysAspAsn*** 72
GTTCTGATGTCATACTGCGATAACTAGTGCGGAAAACTCGACAATTTG 382
ACAATTTTCGTAAACGCCTGATTTATATTAAATAAATTATGTACTAAA 430
ATAATATATGTTTTATTAATGGTACC-3' 456

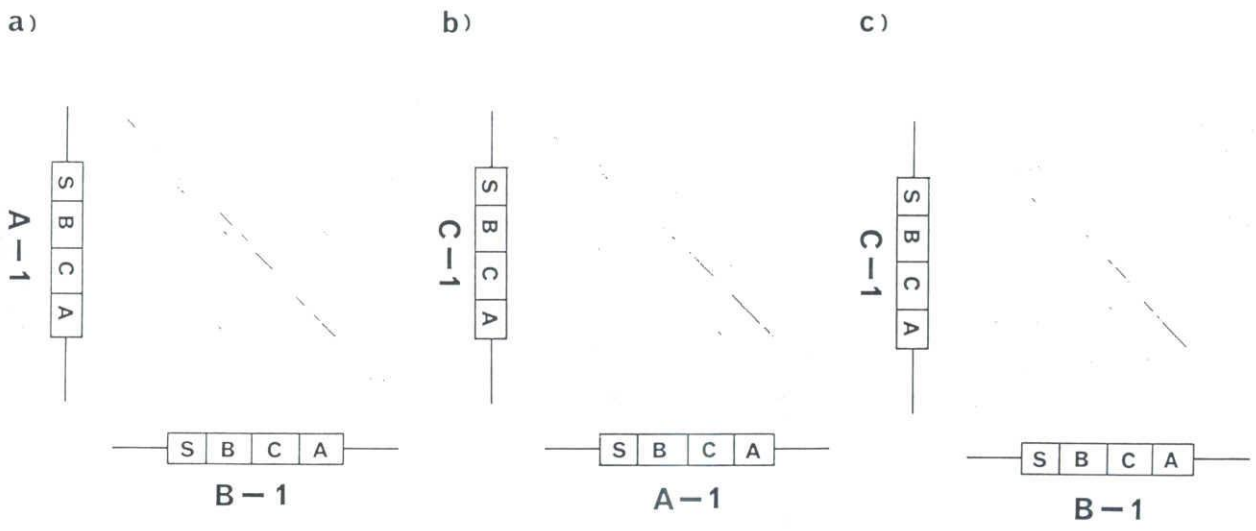
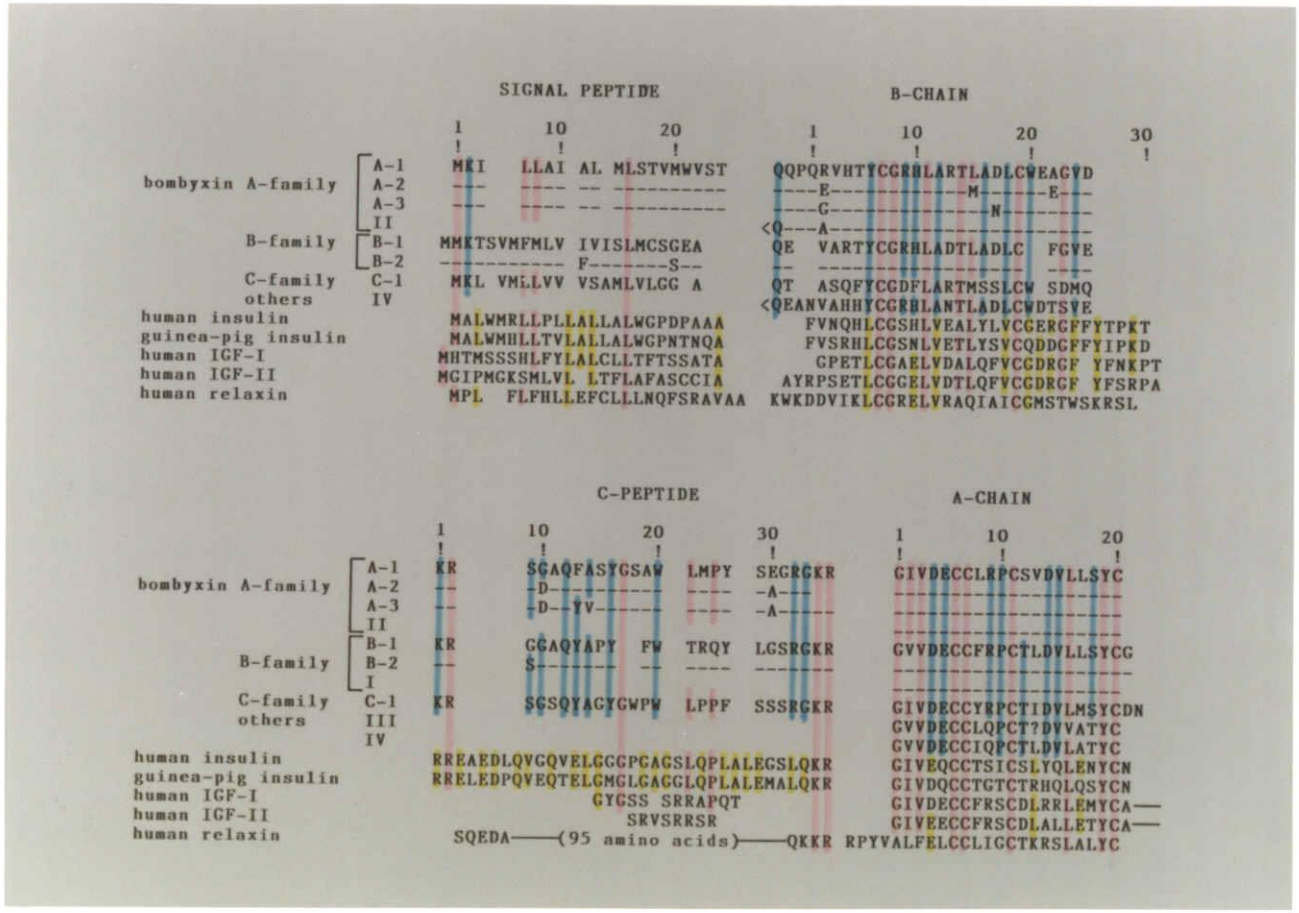
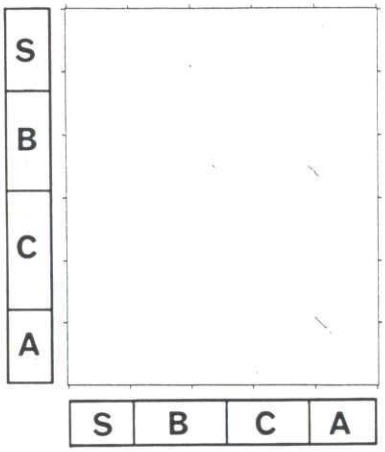

a)
S B C A
A－1
S B C A
B－1
b)
S B C A
C－1
S B C A
A－1
c)
S B C A
C－1
S B C A
B－1

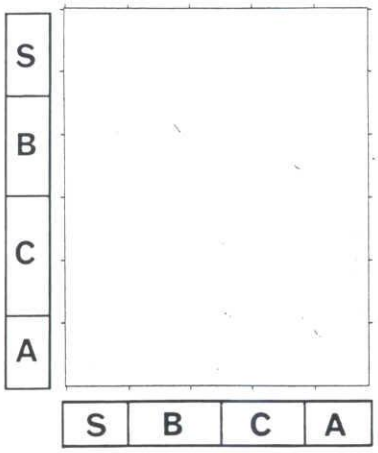
図 9


SIGNAL PEPTIDE
B-CHAIN
C-PEPTIDE
A-CHAIN
bombyxin A-family
A-1
A-2
A-3
II
B-family
B-1
B-2
C-family
C-1
others
IV
I
III
human insulin
guinea-pig insulin
human IGF-I
human IGF-II
human relaxin
SQEDA——(95 amino acids)——QKKR

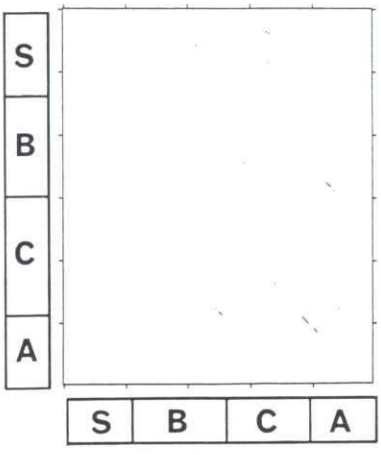
a)




b)




c)



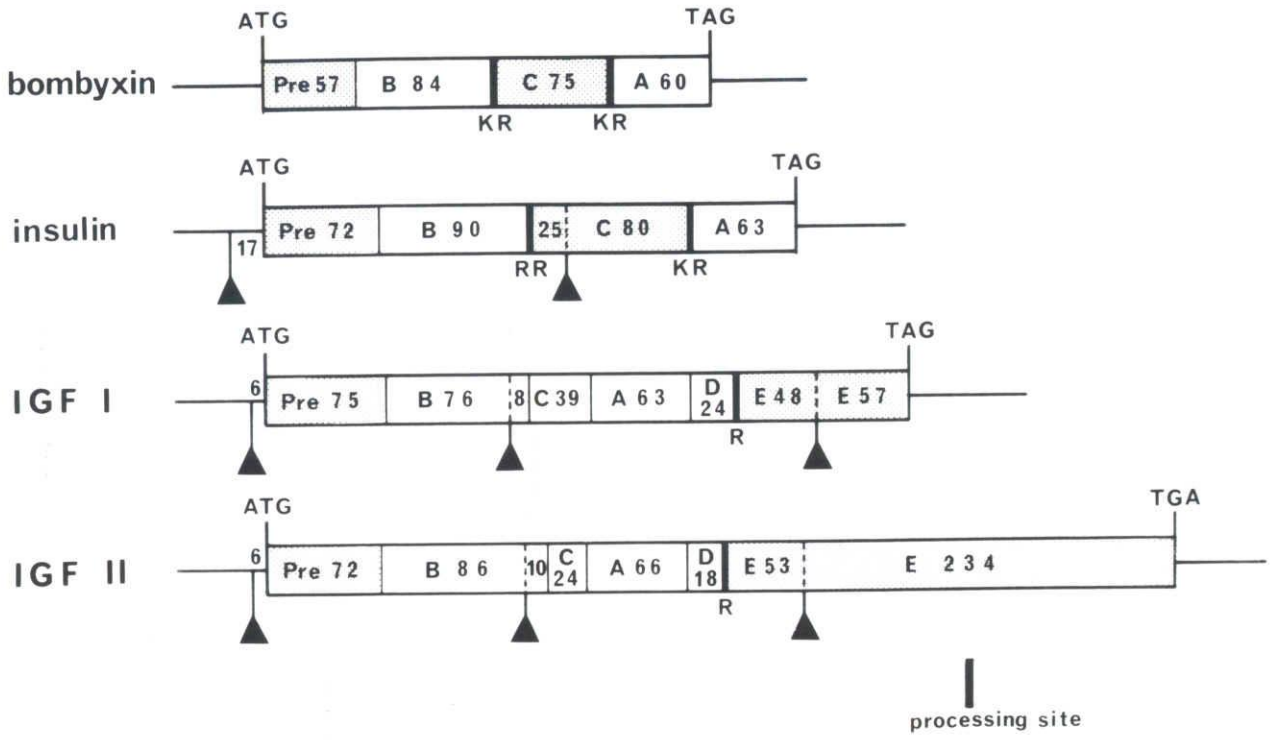
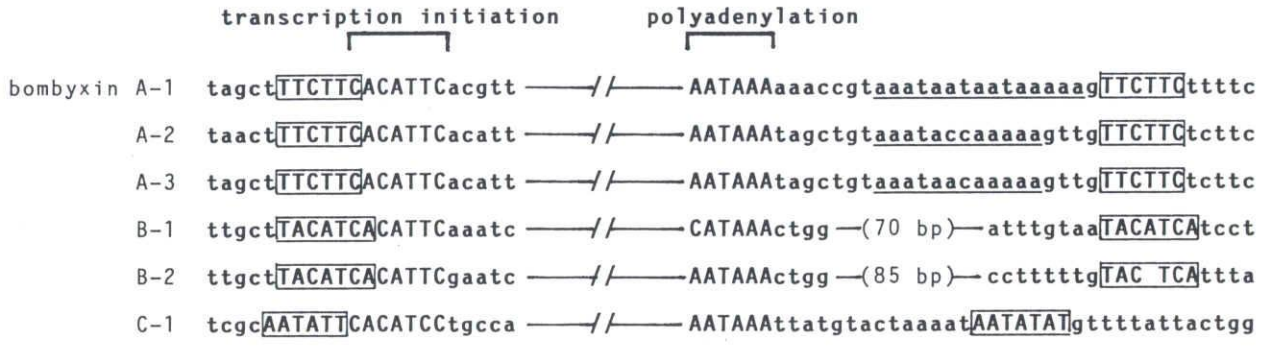

ATG
TAG
bombyxin
Pre 57
B 84
C 75
A 60
KR
KR
ATG
TAG
insulin
17
Pre 72
B 90
25
C 80
A 63
RR
KR
ATG
TAG
IGF I
6
Pre 75
B 76
8
C 39
A 63
D 24
E 48
E 57
R
ATG
TGA
IGF II
6
Pre 72
B 86
10
C 24
A 66
D 18
E 53
E 234
R
processing site

図 1 2


transcription initiation
polyadenylation
bombyxin A-1 tagctTTCTTCACATTCacgtt ——//—— AATAAAaaaccgtaaataataataaaaagTTCTTCttttc
A-2 taactTTCTTCACATTCacatt ——//—— AATAAAtagctgtaaataccaaaaagttgTTCTTCtcttc
A-3 tagctTTCTTCACATTCacatt ——//—— AATAAAtagctgtaaataacaaaaagttgTTCTTCtcttc
B-1 ttgctTACATCACATTCaaatc ——//—— CATAAActgg —(70 bp)— atttgtaaTACATCAtcct
B-2 ttgctTACATCACATTCgaatc ——//—— AATAAActgg —(85 bp)— ccttttgTAC TCAttta
C-1 tcgcAATATTCACATCCtgcca ——//—— AATAAAttatgtactaaaatAATATATgttttattactgg

# 副論文

1. Iwami, M., Kawakami, A., Ishizaki, H., Takahashi, S. Y., Adachi, T., Suzuki, Y., Nagasawa, H. and Suzuki, A., 1989. Cloning of a gene encoding bombyxin, an insulin-like brain secretory peptide of the silkmoth Bombyx mori with prothoracicotropic activity. Develop. Growth Differ., 31, 31-37.

2. Iwami, M., 1990. The gene encoding bombyxin, a brain secretory peptide of Bombyx mori: structure and expression. In Molting and Metamorphosis (Eds. E. Ohnishi and H. Ishizaki), in press. Japan Scientific Societies Press, Tokyo, and Springer-Verlag, Heidelberg.

Develop. Growth & Differ., 31 (1), 31–37 (1989)

Development
Growth & Differentiation

# Cloning of a Gene Encoding Bombyxin, an Insulin-Like Brain Secretory Peptide of the Silkmoth *Bombyx mori* with Prothoracicotropic Activity

(*Bombyx mori*/brain peptide/bombyxin/insulin/IGF)

Masafumi Iwami*[1], Atsushi Kawakami*, Hironori Ishizaki*, Susumu Y. Takahashi*,**, Takashi Adachi*,†, Yoshiaki Suzuki†, Hiromichi Nagasawa†† and Akinori Suzuki††

*Biological Insitute, Faculty of Science, Nagoya University, Nagoya 464-01, Japan;
**College of General Education, Yamaguchi University, Yamaguchi 753, Japan;
†National Institute for Basic Biology, Okazaki 444, Japan;
††Department of Agricultural Chemistry, Faculty of Agriculture, The University of Tokyo, Tokyo 113, Japan.

A genomic DNA encoding bombyxin, a 5kD brain peptide of the silkmoth *Bombyx mori* with prothoracicotropic hormone activity, has been isolated. The nucleotide sequence coding for bombyxin shows high homology with insulin-gene family members and the overall organization of the preprobombyxin gene is the same as in preproinsulin genes, indicating that bombyxin shares a common ancestral molecule with insulin-family peptides. The bombyxin gene has no intron contrasting to other members of insulin-gene family.

## Introduction

Prothoracicotropic hormone (PTTH), a brain secretory peptide of the insect, stimulates the prothoracic glands to synthesize and release ecdysone, the steroid required for insect growth and metamorphosis. The brain of the silkmoth *Bombyx mori* produces, besides its own PTTH (MW. ca. 22,000, previously termed 22K-PTTH, ref. 20), another smaller peptide bombyxin (MW. ca. 5,000, previously termed 4K-PTTH, refs. 20, 21, 30) that shows the prothoracicotropic activity when assayed with the saturniid moth *Samia cynthia ricini* but not with *Bombyx* (19). Bombyxin comprises heterogeneous molecular species and each molecule consists of nonidentical A- and B-chains. Thus, the primary structure has been determined completely for bombyxin-IIs (34, 35) and -IV (35) and partially for bombyxin-I, -III, and -V (23, 33). These molecular forms differ from one another by only minor amino acid substitutions. These bombyxin molecules show considerable homology in the amino acid sequence with insulin-family peptides. The A- and B-chains of bombyxin have been proved to link with disulfide bonds in exactly the same way as in insulin (35) and the insulin-like tertiary structure of bombyxin has been predicted by comprehensive computer analyses (23).

We have recently started to isolate and characterize the gene coding for bombyxin. The usefulness of information on this line is obviously manifold: elucidation of the gene structure may give a deeper insight into evolution of insulin-family peptides; gene expression and its regulation during development may present a clue for finding the function of bombyxin in *Bombyx*; the nature of bombyxin microheterogeneity may be defined in regard to whether due to racial or individual genetic variation in *Bombyx* or due to multicopy genes of bombyxin in the *Bombyx* genome. We report here the cloning and characterization of a bombyxin gene from a *Bombyx* genomic library using synthetic oligonucleotide probes. The results indicate that the overall organization of the preprobombyxin gene is to code for signal peptide/B-chain/C-peptide/A-chain and the nucleotide sequence for A- and B-chains are highly conserved compared to insulin-family genes, corroborating the previous notion derived from peptide analyses that the bombyxin belongs to the insulin family.

[1] All correspondence should be addressed to Masafumi Iwami, Biological Institute, Faculty of Science, Nagoya University, Chikusa-ku, Nagoya 464-01, Japan.

## Materials and Methods

*DNA and RNA preparation* High molecular weight DNA was extracted from the silk gland of *Bombyx mori* (Kinshu × Syowa) by SDS/CsCl gradient method (48). RNA from the silk gland (5th instar) and the larval brain (4–5th instar) of *Bombyx* (J112 × C115) was extracted and purified as described (18). Poly(A)$^+$ RNA was selected by oligo(dT)-cellulose column (2).

*Isolation of the bombyxin genomic DNA* Two kinds of synthetic oligonucleotides designed on the basis of amino acid sequence of bombyxin-II (34) were used to isolate genomic clones for bombyxin. Probe A is a 51-mer antisense oligonucleotide (5′-GCAGTAGCTGAGGAGGACGTCGACGCTGCAGGGTCTGAGGCAGCACTCGTC-3′) corresponding to Asp 4 to Cys 20 of the A-chain which was designed by using the most frequently used codons for 30k-protein of *Bombyx* (37). Probe B (16 mixture), a 51-mer antisense oligonucleotide (5′-CTCCCAGCACAG$^{A}_{G}$TC$^{A}_{G}$GCCAGGGTG-CG$^{A}_{G}$GCCAGGTGGCG$^{G}_{T}$CCGCAGTAGGT-3′) corresponding to Thr 8 to Glu 24 of the B-chain, was designed by the use of the codon frequency of *Drosophila* (27) to decide the nucleotide assignment at redundant positions. A genomic library (*Sau*3AI partially digested, $1.2\times10^5$ independent λEMBL3 clones) of silk gland DNA from *B. mori* was initially screened with radio-labeled probe A and then with probe B. For both screenings with probes A and B, the hybridization solution contained 6× SSC (SSC: 0.15 M NaCl/0.015 M Na-citrate), 10× Denhardt (Denhardt: 0.02% bovine serum albumin/0.02% Ficoll/0.02% polyvinylpyrrolidone), 1% SDS, 500 μg/ml yeast tRNA and 1 pmol/ml $^{32}$P-labeled probe DNA for 20 hr at 52°C. Post-hybridization wash was carried out in 6× SSC/0.1% SDS for 2 hr, and in 2× SSC/0.1% SDS for 30 min at 52°C. Six hybridizing recombinant phages were isolated and four of them were also hybridized with probe B. One of these phages, λ4K112, was further analyzed.

*DNA analysis* The *Hin*dIII fragment of the insert of λ4K112 which hybridized with both probe A and B was subcloned into pUC vectors (49). The deletion mutants for sequencing and those for the bombyxin probe were constructed by the method of exonuclease III digestion (16). Nucleotide sequence was determined on both strands of the *Hin*dIII fragment by the chain-termination method (38) with 7-deaza-dGTP and $^{35}$S-dCTP (31). Nucleotide sequence data was processed with the DNA analysis software (DNASIS, Hitachi software Engineering).

*Northern hybridization analysis* Poly(A)$^+$ RNA (3.3 μg) from the silk gland and 5 μg of poly(A)$^+$ RNA from the larval brain were glyoxalated and electrophoresed on 1% agarose gel (29). The fractionated RNAs were transferred to nylon membrane (45). The hybridization was carried out at stringest condition for 16hr at 42°C in 50% formamide containing 5× SSC, 10× Denhardt, 1% SDS, 300 μg/ml sonicated salmon sperm DNA and the $^{32}$P-labeled probe for the bombyxin coding region (position 505–1045 in Fig. 1). The membrane was washed and then exposed to Kodak XAR-5 film with intensifying screen at −70°C.

## Results

A 1045 base-pair (bp) *Hin*dIII fragment of the insert of λ4k112 which hybridized with both probe A and B has been sequenced to reveal an open reading frame of 276 nucleotides which contained two segments apparently capable of coding for the 20 amino acid residues of the bombyxin A-chain and 28 residues of the B-chain (Fig. 1). All the residues of these two chains but one, position 5 of the B-chain, are identical to those of bombyxin-II (34). Peptide analyses have previously revealed the existence of multiple molecular forms of bombyxin with minor residue substitution, and the position 5 of the B-chain appeared to be a variant site (34). The N-terminal pyroglutamate of bombyxin-II B-chain is most probably generated by the post-translational modification of Gln 1 (46). Northern blot analysis with this gene as a probe indicates that bombyxin mRNA is transcribed in the larval brain (Fig. 2). The length of the mRNA (about 0.6 kb) is consistent with that predicted from this bombyxin gene.

Three potential initiation codons are found at positions −5, −10 and −19. In view of the general features for the signal peptide (47) and the consensus sequence for translation start sites in *Drosophila* (7), we suspect that Met −19 may be the initiator. Flanking the B-chain C-terminus and A-chain N-terminus, there are paired basic amino acid residues (Lys-Arg) suggesting that the post-translational proteolytic cleavage may occur at these sites (8). The fact that the region between the A- and B-chains resembles the C (connecting)-peptide of proinsulins in its length and the net charge of amino acid residues which are essential

```
5'-AAGCTTGTGGAACCTCGCAACAAACGAGACCGTCTTAACCTAAGGCTTCAAAGAATGAT    59
  TTCAACGTTGTAACCCGGTCACCACTTGAAACAAGCAGTTTCTCTATGGTAAAATTTTCC   119
  ACTACCAATCTTTTGAAATGTTTCAAGCTATCAAAGCAATTGGACCACGCCATACTTTCA   179
  CAAATAAATCATAAAACATGATTACTACAGCTCTGATGAAGTCATAAACGCTCGCTTCCA   239
  GCCACGCGTGCATCGAGCGTTATGAACTGGTTCTAAGTTCACCACTGATACCGGATGATG   299
  CTTACTTTGTCTCCTAGTTACTACCTACTACATAATGAAGATGACAAATATTTATAAATT   359
  TTGCTAGGCGTTCAAAGTTAAAAGTTGCGTCAAGTAAATGTACAACATTTAATTCAGTTC   419
  ATTGGTTAATTATTGGAACCTAGGCAATGTCATATAAATCACATGCCCATGTGTTATTAT   479
  ATAAATATTCATTAACTAATCGTCCATATATAAAGGAAGGGTTGCTTAGCTTTCTTCACA   539
                                                  ******
  TTCACGTTTCATCTGTGGCATCAGCGAGACGTCCTCCGAAACACACTTTAGTCGCCAAAC   599

                    SIGNAL PEPTIDE
  -19                -10                 1
  MetLysIleLeuLeuAlaIleAlaLeuMetLeuSerThrValMetTrpValSerThrGln
  ATGAAGATACTCCTTGCTATTGCATTAATGTTGTCAACAGTAATGTGGGTGTCAACACAA   659

                B-CHAIN
  ------------------------------------------------------------
                  10                  20
  GlnProGlnArgValHisThrTyrCysGlyArgHisLeuAlaArgThrLeuAlaAspLeu
  CAGCCACAAAGAGTGCACACGTACTGCGGGCGTCACTTGGCTCGCACTCTGGCCGACCTG   719

                C-PEPTIDE
  ------------------------------------------------------------
                  30                  40
  CysTrpGluAlaGlyValAspLysArgSerGlyAlaGlnPheAlaSerTyrGlySerAla
  TGCTGGGAAGCGGGCGTGGACAAGCGCAGCGGTGCTCAGTTCGCGAGCTACGGCTCCGCG   779

                                A-CHAIN
  ------------------------------------------------------------
                  50                  60
  TrpLeuMetProTyrSerGluGlyArgGlyLysArgGlyIleValAspGluCysCysLeu
  TGGCTGATGCCGTACTCGGAAGGACGCGGCAAACGAGGCATCGTGGATGAGTGTTGTCTC   839

  ------------------------------
                  70
  ArgProCysSerValAspValLeuLeuSerTyrCysAmb
  AGACCCTGCAGCGTGGACGTGCTTCTGTCGTACTGTTAGACCATTCCTTTACCGAATGGT   899
  TCGTTCTCAATACATTTGGAAATAAAAAACCGTAAATAATAATAAAAAGTTCTTCTTTTC   959
                                                  ******
  CTGAACCACAATTGAAAGTGTTGACTGGAGATTTAGATACAGTTTTTTAGACATCATTTA  1019
  GATGATTTTGATTGGCAAGCAAGCTT-3'                                 1045
```

Fig. 1 Nucleotide sequence of the genomic DNA coding for bombyxin. Nucleotides are numbered sequentially from the *Hin*dIII site at the 5′-end. The predicted amino acid sequence shown above is numbered sequentially from the N-terminus of bombyxin-II B-chain. Signal peptide, B-chain, C-peptide and A-chain are indicated by brackets. Promoter-like sequences, CAT box (GGCAATGT) and TATA box (TATATAA), are hatched. Polyadenylation signals (AATAAA) and a downstream 12bp-conserved element (ATACAGTTTTTT) (36) are underlined. Typical transription initiation sequence of insects (ACATTC) (17) is boxed and short-direct repeats (6bp; TTCTTC) are indicated by asterisks under the nucleotides.

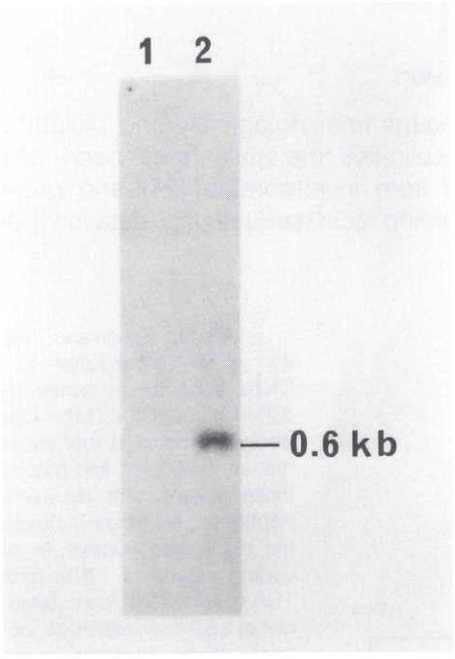



Fig. 2 Northern hybridization of bombyxin mRNA. Lane 1: poly(A)$^+$ RNA (5 $\mu$g) from the silk gland, lane 2: poly(A)$^+$ RNA (3.3 $\mu$g) from 5th-instar larval brain. kb: kilo bases.

to the formation of the proper tertiary structure of these propeptides (9, 42) leads us to presume that this 25-residue segment is equivalent to the C-peptide of proinsulins. Immediately following the A-chain, there is an amber termination codon indicating the absence of D- and E-peptides comparable to those in IGF-I, -II, and their precursors (5, 22). The overall structure of the preprobombyxin gene is thus to code for signal peptide/B-chain/C-peptide/A-chain, being exactly the same as that of the preproinsulin genes (3).

The homology of the deduced amino acid sequence of bombyxin with the sequence of insulin-family peptides is high in the A- and B-chains, particularly in A-chain, while it is less marked in signal and C-peptides (Fig. 3), consistent with the general trend seen among the insulin-family members previously sequenced (42). The homology at the nucleotide sequence level is higher than that at the amino acid sequence level. For example, the nucleotide sequence of the bombyxin gene possesses 57–68% homology in the A-chain and 43–49% in the B-chain with human insulin, IGF-I, and IGF-II. The highest degree of the nucleotide

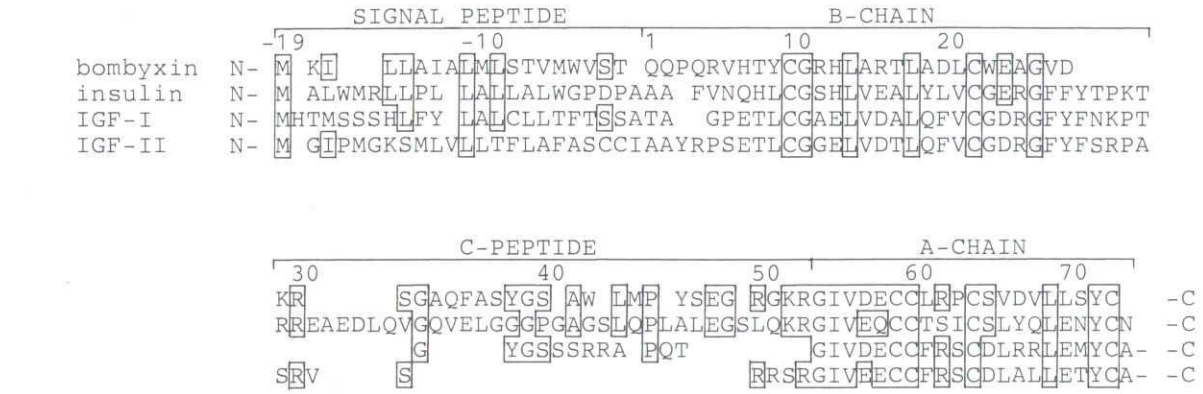



Fig. 3 Amino acid sequence homology between preprobombyxin and three other human insulin-family prepropeptides (3, 5, 22). The boxes indicate identical residues. Amino acids are numbered from the N-terminus of bombyxin-II B-chain.

sequence homology (68%) is seen in the A-chain between bombyxin and IGF-II (5), even higher than between human insulin (3) and IGF-II (63%).

The 5'-flanking region of the gene contains TATA-box like sequence (TATATAA) and CAT-box like sequence (GGCAATGT) 94bp and 158bp upstream from the presumed translation initiation codon, respectively (Fig. 1). Primer extension analysis (data not shown) provided evidence that bombyxin mRNA was transcribed from the typical transcription initiation sequence for insects (ACATTC; 63bp upstream from the translation initiation codon)(17). In the 3'-region, two sets of polyadenylation signal are located 44bp and 64bp downstream from the translation termination codon. A 12bp-conserved element (ATACAG-TTTTTT)(36) is also observed at 120bp downstream from the translation termination codon.

All the insulin-family genes so far characterized have one intron in the 5'-untranslated region. Besides all genes but murine insulin gene I have one or two additional introns in the coding region (12, 26), as schematically shown in Fig. 4 for some representative genes together with the bombyxin gene. By contrast, no splicing signals are found in the bombyxin gene we isolated. Processed genes have poly(A) stretches at the downstream of the polyadenylation signals and short direct repeats flanking the coding region (39). Except for a few functional processed genes (1, 28, 41, 43), processed genes have in-frame stop codons in the coding regions giving rise to pseudogenes. The bombyxin gene characterized here contains 6bp direct repeat (TTCTTC) in the flanking regions (Fig. 1, asterisks) but has neither poly(A) stretches nor in-frame stop codons, indicating that this gene may be functional.

## Discussion

Though immunologically and biologically active insulin-like materials have been amply reported from invertebrates (24) and prokaryotes (25), amino acid sequencing data on bombyxin

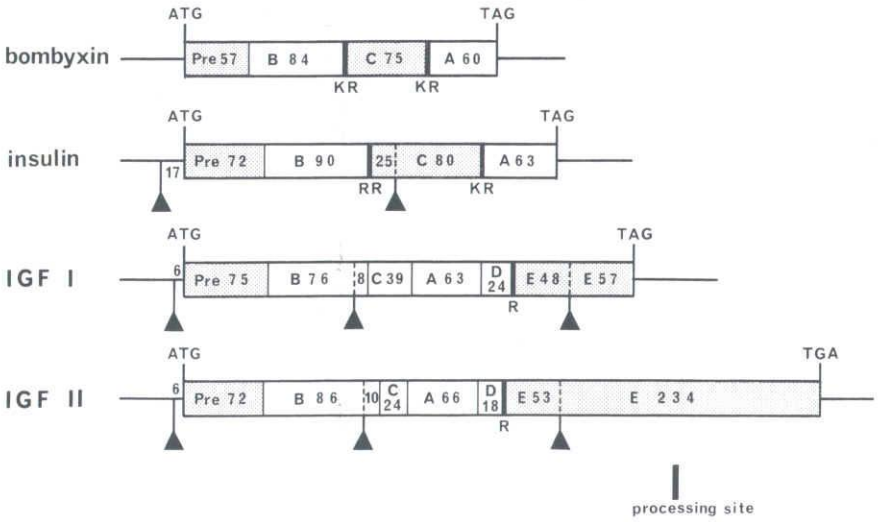



Fig. 4 Schematic representation of the organization of genomic DNAs of bombyxin, human insulin (4), IGF-I (6) and IGF-II (14). Open boxes designate domains that are present in mature molecules and hatched boxes those absent. Pre represents signal peptides. Numerals in boxes indicate the nucleotide number in respective coding segments. ATG and TAG or TGA show the initiation codon and amber or opal termination codon, respectively. Solid triangles represent introns. Paired or single basic amino acid residues by one-letter codes (KR, RR, R) indicate post-translational processing sites.

were the first to demonstrate directly the presence of insulin-related peptide in animals other than vertebrates (33, 34). Recently, in the mollusc *Lymnaea stangalis,* the presence of an insulin-related prepropeptide has been predicted from cDNA analyses and the transcript has been localized by *in situ* hybridization in the cerebral light-green cells, the growth-controlling neuroendocrine cells (40). The present study further demonstrates, on a gene level, that bombyxin belongs to the insulin family, by showing the nucleotide sequence homology and the domain organization of the preprobombyxin gene with the predicted post-translational cleavage site to generate a mature bombyxin. Thus the notion that the invertebrate possesses peptides that share a common ancestral molecule with vertebrate insulin-family peptides is now unequivocally established.

We have determined so far the primary structure of six molecular species of bombyxin completely or partially (23, 33–35) and the purification data indicate that there are still other bombyxin molecules to be sequenced. Accordingly, it is highly probable that there exist multiple bombyxin genes, though it is not known whether they come from genetically different *Bombyx* races or individuals or represent the members of multicopy genes of a *Bombyx* genome. It is interesting to characterize these genes possibly exist and compare these with the one cloned here in regard to its characteristic features, e.g. lack of intron.

Bombyxin failed to stimulate the prothoracic glands of *Bombyx* as far as the pupal-adult (19) and 4th-to-5th larval (44) developements were concerned, and its function in *Bombyx* is still unknown. Of particular interest is the finding that bombyxin is present in the ovary and early developing embryo of *Bombyx* (15) in contrast to the *Bombyx* PTTH (previously termed 22K-PTTH or PTTH-B) which becomes first detectable after the appearance of neuroendocrine system (11). Together with the fact that bombyxin has a high amino acid sequence homology with IGF-II, we suspect that bombyxin might act as a growth factor in embryonic tissues. Insulin has been known to stimulate the growth of imaginal disks and several cell lines of *Drosophila* (13, 32), and the release of triglycerides, diglycerides and fatty acids from fat body in *Agrius cinglulata* and *Danaus plexippus* (10). Furthermore, the brain neurosecretory cells of *Bombyx* have been shown to be immunoreactive to anti-insulin antibody (21, 50), though bombyxin failed to bind to anti-insulin antibody (33). Taken together, it is conceivable that insect nervous tissues produce an array of insulin-like peptides including bombyxin and they play significant roles in various facets of metabolism, cell proliferation, and differentiation, particularly during embryogenic stage.

Acknowledgment: We thank Drs T. Takeda and M. Yamamoto (The Institute of Medical Science, The University of Tokyo) for preparing oligonucleotide probes, and Mr. K. Soma and Miss. I. Kubo for rearing silkworms. This work was supported in part by grants from the Ministry of Education, Science and Culture of Japan (61480020 and for Special Project 61128003), the Ishida Foundation, and the Mitsubishi Foundation.

## References

1. Andersen, R. D., B. W. Birren, S. J. Taplitz and H. R. Herschman, 1986. Rat metallothionein-1 structural gene and three pseudogenes, one of which contains 5′-regulatory sequences. Mol. Cell. Biol., **6**, 302–314.

2. Aviv, H. and P. Leder, 1972. Purification of biologically active globin messenger RNA by chromatography on oligothymidylic acid-cellulose. Proc. Natl. Acad. Sci. USA, **69**, 1408–1412.

3. Bell, G. I., W. F. Swain, R. Pictet B. Cordell, H. M. Goodman and W. J. Rutter, 1979. Nucleotide sequence of a cDNA clone encoding human preproinsulin. Nature, **282**, 525–527.

4. Bell, G. I., R. Pictet, W. Rutter, B. Cordell, E. Tischer and H. M. Goodman, 1980. Sequence of the human insulin gene. Nature, **284**, 26–32.

5. Bell, G. I., J. P. Merryweather, R. Sanchez-Pescador, M. M. Stempien, L. Priestley, J. Scott and L. B. Rall, 1984. Sequence of a cDNA clone encoding human preproinsulin-like growth factor II. Nature, **310**, 775–777.

6. Bell, G. I., D. G. Gerhard, N. M. Fong, R. Sanchez-Pescador and L. B. Rall, 1985. Isolation of the human insulin-like growth factor genes: insulin-like growth factor II and insulin gene are contiguous. Proc. Natl. Acad. Sci. USA, **82**, 6450–6454.

7. Cavener, D. R., 1987. Comparison of the consensus sequence flanking translational start sites in *Drosophila* and vertebrates. Nucl. Acids Res., **15**, 1353–1361.

8. Chan, S. J, S. C. M. Kwok and D. F. Steiner, 1981. The biosynthesis of insulin: some genetic and evolutionary aspects. Diabetes Care, **4**, 4–10.

9. Chan, S. J., S. O. Emdin, S. C. M. Kwok, J. M. Kramer, S. Falkmer and D. F. Steiner, 1981. Messenger RNA sequence and primary structure of preproinsulin in a primitive vertebrate, the atlantic hagfish. J. Biol. Chem., **256**, 7595–7602.

10. Chang, F., 1974. Effects of vertebrate adipokinetic hormones on the rate of in vitro lipid release in insects. Comp. Biochem. Physiol., **49B**, 567–578.

11. Chen, J. H., H. Fugo, M. Nakajima, H. Nagasawa and A. Suzuki, 1987. Neurohormones in developing embryos of the silkworm, *Bombyx mori*: the presence and characteristics of prothoracicotropic hormone B. J. Insect Physiol., **33**, 407–411.

12. Cordell, B., G. Bell, E. Tischer, F. M. DeNoto, A. Ullrich R. Pictet, W. J. Rutter and H. M. Goodman, 1979. Isolation and characterization of a cloned rat insulin gene. Cell, **18**, 533–543.

13. Davis, K. T. and A. Shearn, 1977. In vitro growth of imaginal disks from *Drosophila melanogaster*. Science, **196**, 438–439.

14. Dull, T. J., A. Gray, J. S. Hayflick and A. Ullrich, 1984. Insulin-like growth factor II precursor gene organization in relation to insulin gene family. Nature, **310**, 777–781.

15. Fugo, H., J. Hua, M. Nakajima, N. Nagasawa and A. Suzuki, 1987. Neurohormones in developing embryos of the silkworm, *Bombyx mori*: the presence and characteristics of prothoracicotropic hormone-S. J. Insect Physiol., **33**, 243–248.

16. Henikoff, S., 1984. Unidirectional digestion with exonuclease III creates targeted breakpoints for DNA sequencing. Gene, **28**, 351–359.

17. Hultmark, D., R. Klemenz and W. J. Gehring, 1986. Translational and transcriptional control elements in the untranslated leader of the heat-shock gene *hsp22*. Cell, **44**, 429–438.

18. Ishikawa, E. and Y. Suzuki, 1985. Tissue- and stage-specific expression of sericine genes in the middle silk gland of *Bombyx mori*. Develop. Growth & Differ., **27**, 73–82.

19. Ishizaki, H., A. Mizoguchi, M. Fujishita, H. O'oka, H. Kataoka, A. Isogai, H. Nagasawa, S. Tamura and A. Suzuki, 1983. Species specificity of the insect prothoracicotropic hormone (PTTH): the presence of *Bombyx*- and *Samia*-specific PTTHs in the brain of *Bombyx mori*. Develop. Growth & Differ., **25**, 539–600.

20. Ishizaki, H. and A. Suzuki, 1984. Prothoracicotropic hormone of *Bombyx mori*. *In* Biosynthesis, Metabolism and Mode of Action of Invertebrate Hormones (Eds. J. Hoffmann and M. Porchet), pp. 63–77. Springer-Verlag, Heigelberg.

21. Ishizaki, H., A. Mizoguchi, M. Hatta, A. Suzuki, H. Nagasawa, H. Kataoka, A. Isogai, S. Tamura, M. Fujino and C. Kitada, 1987. Prothoracicotropic hormone (PTTH) of the silkmoth, *Bombyx mori*: 4K-PTTH. *In* Molecular Entomology (Ed. J. H. Law), pp. 119–128. Alan R. Liss, New York.

22. Jansen, M., F. M. A. van Schaik, A. T. Ricker, B. Bullock, D. E. Woods, K. H. Gabbay, A. L. Nussbaum, J. S. Sussenbach and J. L. Van den Brande, 1983. Sequence of cDNA encoding human insulin-like growth factor I precursor. Nature, **306**, 609–611.

23. Jhoti, H., A. N. McLeod, T. L. Blundell, H. Ishizaki, H. Nagasawa and A. Suzuki, 1987. Prothoracicotropic hormone has an insulin-like tertialy structure. FEBS Letters, **219**, 419–425.

24. Kramer, K. J., 1985. Vertebrate hormones in insects. *In* Comprehensive insect physiology, biochemistry, and pharmacology Vol. 7. (Eds. G. A. Kerkut and L. I. Gilbert), pp. 511–536. Pergamon Press, Oxford.

25. LeRoith, D., G. Delahunty, G. L. Wilson, C. T. Roberts Jr., J. Shemer, C. Hart, M. A. Lesniak, J. Shiloach and J. Roth, 1986. Evolutionary aspects of the endocrine and nervous systems. Rec. Progr. Horm. Res., **42**, 549–587.

26. Lomedico, P., N. Rosenthal, A. Efstratiadis, W. Gilbert, R. Koloden and R. Tizard, 1979. The structure and evolution of the two nonallelic rat preproinsulin gene. Cell, **18**, 545–559.

27. Maruyama, T., T. Gojobori, S. Aota and T. Ikemura, 1986. Codon usage tabulated from the GenBank genetic sequence data. Nucl. Acids Res., **14**, r151–r197.

28. McCarrey, J. R. and K. Thomas, 1987. Human testis-specific PGK genes lacks introns and possesses characteristics of a processed gene. Nature, **326**, 501–505.

29. McMaster, G. K. and G. G. Carmichael, 1977. Analysis of single- and double-stranded nucleic acids on polyacrylamide and agarose gels by using glyoxal and acridine orange. Proc. Natl. Acad. Sci. USA, **74**, 4835–4838.

30. Mizoguchi, A., H. Ishizaki, H. Nagasawa, H. Kataoka, A. Isogai, S. Tamura, A. Suzuki, M. Fujino and C. Kitada, 1987. A monoclonal antibody against a synthetic fragment of bombyxin (4K-prothoracicotropic hormone) from the silkmoth, *Bombyx mori*: characterization and immunohistochemistry. Mol. Cell. Endocr., **51**, 227–235.

31. Mizusawa, S., S. Nishimura and F. Seela, 1986. Improvement of the dideoxy chain termination method of DNA sequencing by use of deoxy-7-deazaguanosine triphosphate in place of dGTP. Nucl. Acids Res., **14**, 1319–1324.

32. Mosna, G., 1981. Insulin can completely replace serum in *Drosophila melanogaster* cell cultures in vitro. Experientia, **37**, 466–467.

33. Nagasawa, H., H. Kataoka, A. Isogai, S. Tamura, A. Suzuki, H. Ishizaki, A. Mizoguchi, Y. Fujiwara and A. Suzuki, 1984. Amino-terminal amino acid sequence of the silkworm prothoracicotropic hormone: homology with insulin. Science, **226**, 1344–1345.

34. Nagasawa, H., H. Kataoka, A. Isogai, S. Tamura, A. Suzuki, A. Mizoguchi, Y. Fujiwara, A. Suzuki,, S. Y. Takahashi and H. Ishizaki, 1986. Amino acid sequence of a prothoracicotropic hormone of the silkworm *Bombyxin mori*. Proc. Natl. Acad. Sci. USA, **83**, 5840–5843.

35. Nagasawa, H., K. Maruyama, B. Sato, H. Hietter, H. Kataoka, A. Isogai, S. Tamura, H. Ishizaki, T. Senba and A. Suzuki, 1988. Structure and synthesis of bombyxin from the silkworm, *Bombyx mori*. *In* Peptide Chemistry 1987. (Eds. T. Shiba and S. Sakakibara), pp. 123–126. Protein Research Foundation, Osaka.

36. Renan, M. J., 1987. Conserved 12-bp element downstream from mRNA polyadenylation sites. Gene, **60**, 245–254.

37. Sakai, N., S. Mori, S. Izumi, K. Haino-Fukushima, T. Ogura, H. Maekawa and S. Tomino, 1988. Structures and expression of mRNAs coding for major plasma proteins of *Bombyx mori*. Biochem. Biophys. Acta, **949**, 224–232.

38. Sanger, F., S. Nicklen and A. R. Coulson, 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA, **74**, 5463–5467.

39. Sharp, P. A. 1983. Conversion of RNA to DNA in mammals: *Alu*-like elements and pseudogenes. Nature, **301**, 471–472.

40. Smit, A. B., R. Vreugdenhil, R. H. M. Ebberink, W. P. M. Geraerts, J. Klootwijk and J. Joosse, 1988. Growth-controlling molluscan neurons produce the precursor of an insulin-related peptide. Nature, **331**, 535–538.

41. Stein, J. P., R. P. Munjaal, L. Lagace, E. C. Lai, B. W. O'Malley and A. R. Means, 1983. Tissue-specific expression of a chicken calmodulin pseudogene lacking intervening sequences. Proc. Natl. Acad. Sci. USA, **80**, 6485–6489.

42. Steiner, D. F., S. J. Chan, J. M. Welsh and S. C. M. Kwok, 1985. Structure and evolution of the insulin gene. Ann. Rev. Genet., **19**, 463–484.

43. Stutz, F. and G. Spohr, 1987. A processed gene coding for a sarcomeric actin in *Xenopus laevis* and *Xenopus tropicalis*. EMBO J., **6**, 1989–1995.

44. Suzuki, C. and H. Ishizaki, 1986. Prothoracicotropic hormone bioassay: *Bombyx* larva assay. Int. J. Invert. Reprod. Develop., **10**, 256–274.

45. Thomas, P. S., 1980. Hybridization of denatured RNA and small DNA fragments transferred to nitrocellulose. Proc. Natl. Acad. Sci. USA, **77**, 5201–5205.

46. Turner, A. J., 1986. Processing and metabolism of neuropeptides. *In* Essays in Biochemistry 22. (Eds. R. D. Marchall and K. F. Tipton), pp. 69–119. Academic Press, London.

47. Watson, M. E. E., 1984. Compilation of published signal sequences. Nucl. Acids Res., **12**, 5145–5464.

48. Weeks, D. P., N. Beerman and O. M. Griffith, 1986. A small-scale five-hour procedure for isolating multiple samples of CsCl-puriried DNA: application to isolations from mammalian, insect, higher plant, algal, yeast, and bacterial sources. Anal. Biochem., **152**, 376–385.

49. Yanisch-Perron, C., J. Vieira and J. Messing, 1985. Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. Gene, **33**, 103–119.

50. Yui, R., T. Fujita and S. Ito, 1980. Insulin-, gastrin-, pancreatic polypeptide-like immunoreactive neurons in the brain of the silkworm, *Bombyx mori*. Biomed. Res., **1**, 42–46.

*(Received July 7, 1988; accepted July 28, 1988)*

Note added.

The nucleotide sequence data reported in this paper will appear in the EMBL, GenBank and DDBJ Nucleotide sequence Databases under the accession number d00340.

In *Molting and Metamorphosis* (Eds. Ohnishi, E. and Ishizaki, H), in press (1990). Japan Scientific Societies Press, Tokyo, and Springer-Verlag, Heidelberg.

# The Gene Encoding Bombyxin, a Brain Secretory Peptide of *Bombyx mori*: Structure and Expression

Masafumi Iwami

Biological Institute, Faculty of Science, Nagoya University, Nagoya 464-01, Japan

Number of Figures: 8; Number of Table: 1

Running Head: Bombyxin Genes

## Introduction

Prothoracicotropic hormone (PTTH), a brain secretory peptide of insects, stimulates the prothoracic glands to synthesize and release ecdysone, the steroid required for insect growth and metamorphosis (Ishizaki and Suzuki, 1984; Bollenbacher and Granger, 1985). In addition to its own PTTH (MW. ca. 30,000-Da, ref. Kataoka *et al.*, 1990), the brain of the silkmoth *Bombyx mori* produces a 5,000-Da peptide bombyxin that manifests the prothoracicotropic activity when tested with the saturniid moth *Samia cynthia ricini* (Ishizaki *et al.*, 1983; Nagasawa *et al.*, 1984a).

Bombyxin represents a peptide group of heterogeneous molecular species that differ by only a few amino acid substitutions. Bombyxin molecules are heterodimers consisting of A- and B-chains whose amino acid sequences show considerable homology with vertebrate insulin-family peptides (Nagasawa *et al.*, 1984b; 1986). The primary structure of bombyxin has been determined completely for bombyxin-IIs (Nagasawa *et al.*, 1986; 1988) and -IV (Maruyama *et al.*, 1988) and partially for bombyxin-I, -III, and -V (Nagasawa *et al.*, 1984b; Jhoti *et al.*, 1987). The A- and B-chains of bombyxin have been proved to link with disulfide bonds in exactly the same way as in insulin (Nagasawa *et al.*, 1988) and the insulin-like tertiary structure of bombyxin has been predicted by computer analyses based on the assumption of homology with porcine insulin (Jhoti *et al.*, 1987).

We have recently started studies using recombinant DNA technology to clone the bombyxin cDNA and genomic DNA (Iwami *et al.*, 1989; 1990; Adachi *et al.*, 1989; Kawakami *et al.*, 1989). Elucidation of the gene struc-

ture may give a deeper insight into not only the evolution, gene expression, and its regulation of bombyxin, but also molecular understanding of insulin-family peptides in general. I review here the organization, structure and expression of bombyxin genes and speculate the physiological function of bombyxins. The bombyxin genes have unique features in their organization and structure, such as the presence of multi-copy genes, formation of gene-pairs, and characteristics as the functional processed gene. These facts suggest that different mechanisms for the evolution of insulin-family genes have been operating in invertebrates and vertebrates.

## Isolation of Bombyxin Genes

Two kinds of oligonucleotide probes designed on the basis of the amino acid sequence of bombyxin-II (Nagasawa *et al.*, 1986) have been used to isolate the cDNA and gene encoding bombyxin. Two clones, λBb360 and λBb204, from a *Bombyx* larval brain cDNA library (Adachi *et al.*, 1989) and two clones, λ4K105 and λ4K112, from a *Bombyx* genomic library (Iwami *et al.*, 1989; Kawakami *et al.*, 1989) were characterized. One of the genomic clones (λ4K105) contained four bombyxin genes and the other (λ4K112) contained one bombyxin gene. Five bombyxin genes from the genomic library have been classified into family A and B according to their sequence homology (Kawakami *et al.*, 1989). Thus, bombyxin genes A-1, A-2, A-3, B-1, and B-2 have been defined. The nucleotide sequence of the λBb360 cDNA completely matched the corresponding region of gene A-2. Thus, the λBb360 cDNA has apparently been derived from the gene A-2 transcript. The λBb204 cDNA encoded a bombyxin whose deduced amino acid sequence was apparently different from other bombyxins previously characterized, but this λBb204 cDNA was of partial length. The complete gene encoding this new bombyxin was cloned using the λBb204 cDNA as a probe and named gene C-1 (Iwami *et al.*, 1990). In this way, six genes which encode different bombyxin molecules have been characterized.

## Organization of Bombyxin Genes

Four bombyxin genes (A-2, A-3, B-1, and B-2) form a cluster in a 9-

kilo base-pair (kb) *Bombyx* genomic DNA segment (Figure 1). Two genes which belong to different families are closely apposed to form a pair (A-2/B-1 or A-3/B-2) with opposite transcriptional orientation, presumably functioning as a regulatory unit for transcription (Kawakami *et al.*, 1989). Genes B-1 and A-2 are about 1.1 kb apart and genes B-2 and A-3 are separated by about 2.3 kb. The wider space between B-2 and A-3 is due to the presence of an insertion sequence (BIS). Gene pairs B-1/A-2 and B-2/A-3 form a highly homologous tandem repeat, suggesting that these gene pairs must have been generated by duplication of an original gene pair. Genes A-1 and C-1 also form pairs with the B-family genes (Kondo *et al.*, unpublished results). A similar gene-pair organization has been extensively studied for the moth chorion protein superfamily (Goldsmith and Kafatos, 1984; Kafatos *et al.*, 1987).

The copy number of bombyxin family A- and B-genes was estimated by genomic Southern hybridization to be 12 and 10, respectively (Kawakami *et al.*, 1989). In addition, the hybridization patterns have suggested that the family A- and B- genes are localized close to one another, possibly forming pairs like the B-1/A-2 and B-2/A-3 gene pairs, and that most of the bombyxin gene copies are clustered in three regions in the genome. Like families A and B, a group of genes closely related to gene C-1 is also present in the *Bombyx* genome (Iwami *et al.*, 1990). Gene C-1 is thus a member of a multigene family, the bombyxin C-family. Furthermore, the data of peptide analyses of native bombyxins (Nagasawa *et al.*, 1984b; 1986; 1988; Jhoti *et al.*, 1987; Maruyama *et al.*, 1988) which showed the presence of bombyxins that could not be assigned to any of the families A, B, and C suggest the

presence of still other bombyxin families to be identified. The existence of multiple bombyxin gene copies per *Bombyx* genome is in sharp contrast to vertebrate insulin-family genes that exist in a single or two copies per haploid genome (Steiner *et al.*, 1985). The difference in the gene copy number and the unique organization of multiple gene copies in the genome suggest that evolutionary mechanism for insulin-family genes of invertebrates differ significantly from that of vertebrates.

## Structure of Bombyxin Genes

The structure of bombyxin A-, B-, and C-family genes is shown schematically in Figure 2, together with that of human insulin gene as a representative of vertebrate insulins. All bombyxin genes encode preprobombyxin, the precursor molecule for bombyxin, with the domain organization of signal peptide/B-chain/proteolytic cleavage signal/C-peptide/proteolytic cleavage signal/A-chain. The structure of preprobombyxins is the same as that of preproinsulin, suggesting that preprobombyxins are posttranslationally modified to form mature bombyxins through the excision of the C-peptide, as in preproinsulin (Chan *et al.*, 1981). The presence of pyroglutamate at the B-chain N-terminus of native bombyxins (Nagasawa *et al.*, 1986) indicates that the B-chain N-terminal glutamine residue of probombyxins is probably converted to pyroglutamate by a posttranslational modification (Turner, 1986).

The homology of the nucleotide sequence of bombyxins with the sequences of insulin-family genes is high in the A-chain region, while it is less marked in other domains. For example, the nucleotide sequence of the A-

chain of the bombyxin A-1 gene possesses 57-68% homology with the corresponding sequences of human insulin, insulin-like growth factor (IGF)-I, and IGF-II genes (Iwami *et al.*, 1989). The highest degree of the nucleotide sequence homology in the A-chain region (68%) is seen between bombyxin A-1 gene and IGF-II gene, being even higher than between human insulin and IGF-II genes (63%).

Homology dot-matrix analysis between bombyxin genes (Iwami *et al.*, 1990) indicates that gene C-1 is related more distantly to the A- and B-family genes than the A- and B-families are related to each other, especially for the B-chain region. This result indicates that the C-family gene is evolutionarily remote from the A- and B-family genes. Table 1 shows the evolutionary distance per nucleotide site calculated using 3ST model of Kimura (1981). In contrast to the A-chain region of bombyxins which is evolutionarily conserved throughout the bombyxin genes, the B-chain region is well conserved only among the A- and B-families. The conservation of the A-chain among bombyxins must be due to a low acceptance of mutation as associated with the structural requirements necessary for maintaining a common physiological function. The high substitution rate in the B-chain of the bombyxin C-1 gene compared to the A- and B-family genes may indicate that the C-1 B-chain has escaped from the stringent structural requirements for the function common to the A- and B-family B-chains. Then the function of C family bombyxins which the B-chain structure contributes to would differ from that of other bombyxin families.

All the bombyxin genes characterized apparently lack introns in the untranslated region and the C-peptide region where insulin genes have introns

except the murine insulin-I gene (Steiner *et al.*, 1985). The murine insulin-I gene has an intron in the C-peptide region, but has not in the untranslated region. Although bombyxin genes lack introns completely, they have the typical eukaryotic transcription promoter (TATA-AAA), transcription initiator (CACATTC), and polyadenylation signal sequences, suggesting that bombyxin genes are transcribed by RNA polymerase II.

The processed gene (Vanin, 1985) or retroposon (Rogers, 1983) has been defined as the gene that was generated by the reverse transcription of a processed mRNA and subsequent reinsertion into the genome. The hallmarks of processed genes are 1) the lack of introns, 2) presence of poly(A) tract 3' to the transcription terminator, and 3) presence of direct repeats bounding the transcribed region. Processed genes in most cases have in-frame stop codons in the coding regions giving rise to pseudogenes, but there exist a few functional processed genes (Andersen *et al.*, 1986; McCarrey and Thomas, 1987; Stutz and Spohr, 1987). All bombyxin genes characterized lack intron and, as shown in Figure 3, have boundary short direct repeats though the repeat is incomplete for B-2 and C-1. In addition, the A-family genes contain the remnants of poly(A) tract immediately preceding the downstream repeat. The structural features of bombyxin genes thus indicate that bombyxin genes are the processed genes which are functional. It is recalled that the murine insulin-I genes have also been inferred to be the functional processed genes (Soares *et al.*, 1985).

## Amino Acid Sequence of Bombyxins

The amino acid sequences of preprobombyxins are aligned with those of native bombyxins, human and guinea-pig insulins, IGF-I, -II, and human relaxin (Figure 4). When the amino acid sequences of bombyxins are compared with those of insulins, striking similarities are found between the residues that contribute to the tertiary structure of insulin and the corresponding residues of bombyxins, as has previously been pointed out by comparing insulin mainly with bombyxin-II (Jhoti *et al.*, 1987). Thus, all residues known to make up the hydrophobic core of insulin, A2, A6, A11, A16, A20, B11, B15, and B19, are all conserved or conserved as hydrophobic in bombyxins. The residues on the insulin protomer surface that contribute to the hydrophobic core, A3, A19, B6, B12, B18, and B24, are likewise conserved or conserved as hydrophobic. A1 and B8 glycines which serve for the maintenance of the insulin conformation through its contribution to the main-chain turn are also conserved. In addition, the hydropathy profiles of preprobombyxins and preproinsulin are also similar to each other, except for the B-chain C-terminal regions (Figure 5). The B-chain C-terminal region of insulin contributes to the hydrophilic surface of the protomer but not to the core important for the insulin conformation. Thus, it appears highly probable that all bombyxins have an insulin-like globular structure.

The amino acid sequences of the A- and B-chains show high homology between the bombyxin A- and B-families. However, the amino acid sequence of C-1 is considerably different from that of other bombyxins. With respect to the B-chain in particular, the difference between bombyxin C-1 and other

bombyxins is prominent. The amino acid sequence homology of bombyxin C-1 with A- and B-family bombyxins (44% and 36%) is significantly lower than the homology between the A- and B-families (61%). Remarkably, four residues at position 9, 10, 16, and 24 of the B-chain are substituted when compared to the A- and B-families in which these four sites are completely conserved (Figure 4). The residues at these sites of C-1 bombyxin also differ from those of insulin. The difference of C-1 from bombyxin-III and -IV is also prominent. Thus, the unique structure of C-family bombyxin suggest its possible unique function, as was argued in a previous section from the nucleotide sequence comparison.

In insulins, B16 and B24 have been implicated in the receptor binding and B9 and B10 are situated close to the proposed receptor-binding region (Pullen *et al.*, 1976). These four surface residues of bombyxin C-1 differ from those of other bombyxins, although these residues are completely conserved in the latter. Assuming that bombyxins have a structure-function relationship analogous to insulin with respect to receptor binding, bombyxin C-1 might differ in its physiological function from other bombyxins by using a different receptor (Iwami *et al.*, 1990). Interestingly no bombyxins structurally related to bombyxin C-1 have yet been isolated from the *Bombyx* head by the purification which was based on the bioassay of the prothoracicotropic activity on *Samia*.

All insulins that form 2-zinc hexamers have histidine at B10 which binds zinc but the hystricomorph and hagfish insulins lack B10 histidine and do not form 2-zinc hexamer (Blundell *et al.*, 1970; Chan *et al.*, 1981; 1984). Bombyxin C-1 also lacks histidine at B10, suggesting that C-1 bombyxin does

not form 2-Zinc hexamer. It has been argued that bombyxins may not form either dimer or hexamer and the presence of histidine at B10 in bombyxin A- and B-families may be a result of convergent evolution (Jhoti *et al.*, 1987).

## Expression of Bombyxin Genes

Northern hybridization experiments demonstrated that brain was the only tissue examined which contained a transcript (ca. 600 nucleotides) as visualized by the bombyxin A- and B-family probes (Kawakami *et al.*, 1989). The localization of the bombyxin mRNA examined by *in situ* hybridization indicates that four pairs of mid-dorsal neurosecretory cells produce the A- and B-family transcripts (Figure 6). These same cells were immunoreactive to a monoclonal antibody against the synthetic bombyxin-I A-chain fragment (Mizoguchi *et al.*, 1987).

Figure 7 shows the developmental change in the bombyxin A-, B-, and C-family mRNA amounts during larval-pupal development of *Bombyx* (Adachi *et al.*, unpublished results). It has been described previously that the A-family mRNA amount per brain total RNA decreased gradually with the growth in the 5th-instar (Adachi *et al.*, 1989). The B- and C-family mRNA amounts also decreased gradually during the development. When these curves are redrawn on the basis of the number of mRNA molecules per brain, however, the titer remains essentially unchanged (3-6 molecules per brain for the A- and C-families; 8-12 molecules per brain for the B-family). Though the amount of bombyxin in the *Bombyx* brain rose significantly at around pupation (Ishizaki, 1969; Ishizaki *et al.*, 1983), the bombyxin mRNA titers per

brain did not change appreciably at that time. Furthermore, the amount of bombyxin per brain (Ishizaki and Suzuki, 1984) is several thousand times larger on the molar basis than its mRNA. Taken together, it seems probable that the posttranscriptional processes are important in regulating the bombyxin titer rather than the transcriptional process. However, this inference is not conclusive because the amount of bombyxin mRNA may have been underestimated in view of the possible presence of mRNAs transcribed by genes of other families.

## Speculation on the Function of Bombyxins

The physiological function of bombyxin has remained unclarified for *Bombyx*, but its insulin-like structure is thought to suggest some essential function in *Bombyx*. Insulin and IGFs control cell proliferation and differentiation in a variety of tissues and at various developmental stages of vertebrates (Froesch *et al.*, 1985). In insect, exogenous insulin has been shown to affect the cell growth as evidenced by the stimulus on the imaginal disks and several cell lines of *Drosophila* (Davis and Shearn, 1977; Mosna, 1981). Furthermore, bombyxins show a high amino acid sequence homology with IGFs which have potent mitogenic (Zapf *et al.*, 1981; Mercola and Stiles, 1988) and morphogenetic (Beck *et al.*, 1987; Stylianopoulou *et al.*, 1988) activities. Taken together, it seems probable that bombyxin possesses the activity as a growth factor of exerting effects on cell growth, proliferation, and differentiation of various tissues and at various developmental stages.

Bombyxin has been reported to accumulate in the abdomen of the

female developing adult of *Bombyx* (Ishizaki, 1969). Fugo *et al.* (1987) have shown that bombyxin is present in the ovary and early developing embryo of *Bombyx*. Our unpublished results (Kawakami *et al.*) showed that the transcript of C-family bombyxin genes was detected in the developing embryo whereas the A- and B-family genes were expressed only at the postembryonic stages. It is recalled that in vertebrates insulin and IGFs exert effects on oocyte maturation (El-Etr *et al.*, 1979) and embryogenesis (Froesch *et al.*, 1985). From these facts, it is tempting to hypothesize that bombyxins, at least those of a particular family, control oogenesis and embryogenesis of *Bombyx*.

## Speculation on the Evolution of Bombyxin Genes

From the nucleotide sequence comparison between bombyxin families, I propose a hypothesis for the bombyxin gene evolution in *Bombyx* as follows (Figure 8). First, a proto-bombyxin gene diverged from the proto-insulin line by gene duplication, then three (or more) bombyxin genes representing the original genes for the A-, B-, and C-family (or more families) genes presently exist diverged from the proto-bombyxin gene, and lastly the three (or more) bombyxin genes were duplicated to give rise to multi-gene families. Short direct repeats bounding the coding region of bombyxin genes are highly conserved both in the sequence and the position within respective families whereas they are largely divergent between the families (Figure 3). Thus, the generation of the processed genes may have occurred after the proto-bombyxin genes of the respective families had diverged. Alternatively, the processed genes may have repeatedly generated from a single proto-bombyxin gene to

settle at different sites of the genome. The settled bombyxin genes then evolved the respective families. The brain neurosecretory cells of *Bombyx* have been shown to contain a molecule immunoreactive to anti-insulin antibody (Yui *et al.*, 1980; Ishizaki *et al.*, 1987), though bombyxin failed to bind to anti-insulin antibody (Nagasawa *et al.*, 1984b). Therefore, it seems likely that the *Bombyx* brain contains an insulin-related peptide structurally and possibly functionally distinct from bombyxins. I presume therefore that the proto-insulin line evolved "insulin", the insulin-related molecule which reacts with the anti-insulin antibody, independently of bombyxins.

## Conclusion

Progress has been made toward the molecular understanding of bombyxin. Our recent studies on bombyxin cDNA and gene structure have revealed a striking similarity in the mode of biosynthesis between bombyxin and insulin. Bombyxin has been presumed to be synthesized first as a large precursor molecule from which the mature bombyxin is posttranslationally generated after the excision of C-peptide, in the same way as in insulin. Thus the conclusion that bombyxin shares a common ancestral molecule with vertebrate insulin-family peptides as first proposed by peptide analysis of purified bombyxins (Nagasawa *et al.*, 1984b; 1986) was substantiated and further extended based on the genetic information. Moreover, the presence of multiple bombyxin gene copies and their organization in the genome have suggested that evolutionary mechanism for insulin-family genes differs between vertebrates and invertebrates.

Recently, in the mollusc *Lymnaea stangalis*, the presence of an insulin-related prepropeptide has been predicted from cDNA analyses and the transcript has been localized by *in situ* hybridization in the growth-controlling neuroendocrine cells (Smit *et al.*, 1988). Now in our laboratory, the structure and organization of as many as thirty bombyxin gene copies have been clarified (Kondo *et al.*, unpublished results). Six clustered genes encoding bombyxin-related prepropeptides have also been isolated from the *Samia* genome (Kimura *et al.*, unpublished results). Furthermore, cDNA encoding an insulin-related prepropeptide has been cloned from a library prepared from the *Locusta* pars intercerebralis (Lageux and Hoffmann, personal communication).

The structural information about these insulin-family genes of invertebrates will contribute to understanding the function and evolution of the insulin-family genes not only of invertebrates but also of vertebrates, and a new facet for the insulin-family peptide evolution will be generated.

**Acknowledgment**

The author thanks Prof. H. Ishizaki for helpful discussion and critical reading of the manuscript, T. Adachi, A. Kawakami, H. Kondo, M. Kimura, and M. Ikeno for making data available for this publication, I. Kubo and K. Soma for technical assistance, Dr. M. Lageux and Prof. J. Hoffmann (Université Louis Pasteur) for providing me with unpublished results. The work presented in this review was supported by grants from the Ministry of Education, Science and Culture of Japan, from the Ishida foundation, and from the Mitsubishi foundation.

## References


Adachi T., Takiya S., Suzuki Y., Iwami M., Kawakami A., Takahashi S. Y., Ishizaki H., Nagasawa H. and Suzuki A. (1989) *J. Biol. Chem.*, **264**, 7681-7685.

Andersen R. D., Birren B. W., Taplitz S. J. and Herschman H. R. (1986) *Mol. Cell. Biol.*, **6**, 302-314.

Beck F., Samani N. J., Penschow J. D., Thorley B., Tregear G. W. and Coghlan J. P. (1987) *Development*, **101**, 175-184.

Bell G. I., Pictet R., Rutter W., Cordell B., Tischer E. and Goodman H. M. (1980) *Nature*, **284**, 26-32.

Bell G. I., Merryweather J. P., Sanchez-Pescador R., Stempien M., Priestley L., Scott J. and Rall L. B. (1984) *Nature*, **310**, 775-777.

Blundell T. L., Dodson G. G., Dodson E., Hodgkin D. C. and Vijayan M. (1970) *Rec. Progr. Horm. Res.*, **27**, 1-40.

Bollenbacher W. E. and Granger N. A. (1985) In *Comprehensive Insect Physiology, Biochemistry and Pharmacology* (Ed. by Kerkut G. A. and Gilbert L. I.), Vol.7, pp.109-151. Pergamon Press, Oxford.

Chan S. J., Kwok S. C. M. and Steiner D. F. (1981) *Diabetes Care*, **4**, 4-10.

Chan S. J., Episkopou V., Zeitlin S., Karathanasis S. K., MacKrell A., Steiner D. F. and Efstratiadis A. (1984) *Proc. Natl. Acad. Sci. USA.*, **81**, 5046-5050.

Davis K. T. and Shearn A. (1977) *Science*, **196**, 438-439.

El-Etr M., Schorderet-Slatkine S. and Baulieu E.-E. (1979) *Science*, **205**, 1397-1399.

Froesch E. R., Schmid C., Schwander J. and Zapf J. (1985) *Ann. Rev. Physiol.*, **47**, 443-467.

Fugo H., Hua J., Nakajima M., Nagasawa H. and Suzuki A. (1987) *J. Insect Physiol.*, **33**, 243-248.

Goldsmith M. R. and Kafatos F. C. (1984) *Ann. Rev. Genet.*, **18**, 443-487.

Hudson P., Haley J., John M., Cronk M., Crawford R., Haralambidis J., Tregear G., Shine J. and Niall H. (1983) *Nature*, **301**, 628-631.

Ishizaki H. (1969) *Dev. Growth Differ.*, **11**, 1-7.

Ishizaki H., Suzuki A., Moriya I., Mizoguchi A., Fujishita M., O'oka H., Kataoka H., Isogai A., Nagasawa H. and Suzuki A. (1983) *Dev. Growth Differ.*, **25**, 585-592.

Ishizaki H. and Suzuki A. (1984) In *Biosynthesis, Metabolism and Mode of Action of Invertebrate Hormones* (Ed. by Hoffman, J. and Porchet, M), pp.63-77. Springer-Verlag, Heidelberg.

Ishizaki H., Mizoguchi A., Hatta M., Suzuki A., Nagasawa H., Kataoka H., Isogai A., Tamura S., Fujino M. and Kitada C. (1987) In *Molecular Entomology* (Ed. by Law J. H.), pp.119-128. Alan R. Liss, New York.

Iwami M., Kawakami A., Ishizaki H., Takahashi S. Y., Adachi T., Suzuki Y., Nagasawa H. and Suzuki A. (1989) *Dev. Growth Differ.*, **31**, 31-37.

Iwami M., Adachi T., Kondo H., Kawakami A., Suzuki Y., Nagasawa H., Suzuki A. and Ishizaki H. (1990) *Insect Biochem.*, in press.

Jansen M., van Schaik F. M. A., Ricker A. T., Bullock B., Woods D. E., Gabbay K. H., Nussbaum A. L., Sussenbach J. S. and Van den Brande J. L. (1983) *Nature*, **306**, 609-611.

Jhoti H., McLeod A. N., Blundell T. L., Ishizaki H., Nagasawa H. and Suzuki A. (1987) *FEBS Lett.*, **219**, 419-425.

Kafatos F. C., Mitsialis S. A., Nguyen H. T., Spoerel N., Tsitilou S. G. and Mazur G. D. (1987) In *Development as an Evolutionary Process* (Ed. by Raff R. A. and Raff E. C.), pp. 161-178. Alan R. Liss, New York.

Kataoka H., Nagasawa H., Isogai A., Ishizaki H. and Suzuki A. (1990) in preparation.

Kawakami A., Iwami M., Nagasawa H., Suzuki A. and Ishizaki H. (1989) *Proc. Natl. Acad. Sci. USA.*, **86**, in press.

Kimura M. (1981) *Proc. Natl. Acad. Sci. USA.*, **78**, 454-458.

Kyte, J. and Doolittle, R. F. (1982) *J. Mol. Biol.*, **157**, 105-132.

Lomedico P., Rosenthal N., Efstratiadis A., Gilbert W., Kolodner R. and Tizard R. (1979) *Cell*, **18**, 545-558.

Maruyama K., Hietter H., Nagasawa H., Isogai A., Tamura S., Suzuki A. and Ishizaki H. (1988) *Agric. Biol. Chem.*, **52**, 3035-3041.

McCarrey J. R., and Thomas K. (1987) *Nature*, **326**, 501-505.

Mercola M. and Stiles C. D. (1988) *Development*, **102**, 451-460.

Mizoguchi A., Ishizaki H., Nagasawa H., Kataoka H., Isogai A., Tamura S., Suzuki A., Fujino M. and Kitada C. (1987) *Mol. Cell. Endocr.*, **51**, 227-235.

Mosna G. (1981) *Experientia*, **37**, 466-467.

Nagasawa H., Kataoka H., Hori Y., Isogai A., Tamura S., Suzuki A., Guo F., Zhong X., Mizoguchi A., Fujishita M., Takahashi S. Y., Ohnishi E. and Ishizaki, H. (1984a) *Gen. Comp. Endocr.*, **53**, 143-152.

Nagasawa H., Kataoka H., Isogai A., Tamura S., Suzuki A., Ishizaki H., Mizoguchi A., Fujiwara Y. and Suzuki A. (1984b) *Science*, **226**, 1344-1345.

Nagasawa H., Kataoka H., Isogai A., Tamura S., Suzuki A., Mizoguchi A., Fujiwara Y., Suzuki A., Takahashi S. Y. and Ishizaki H. (1986) *Proc. Natl. Acad. Sci. USA.*, **83**, 5840-5843.

Nagasawa H., Maruyama K., Sato B., Hietter H., Kataoka H., Isogai A., Tamura S., Ishizaki H., Senba T. and Suzuki A. (1988) In *Peptide Chemistry 1987* (Ed. Shiba, T. and Sakakibara, S.) pp.123-126, Protein Research Foundation, Osaka.

Pullen R. A., Lindsay D. G., Wood S. P., Tickle I. J., Blundell T. L., Wollmer A., Krail G., Brandenburg D., Zahn H., Gliemann J. and Grammeltoft S. (1976) *Nature*, **259**, 369-373.

Rogers J. (1983) *Nature*, **301**, 460.

Smit A. B., Vreugdenhil R., Ebberink R. H. M., Geraerts W. P. M., Klootwijk J. and Joosse J. (1988) *Nature*, **331**, 535-538.

Soares M. B., Schon E., Henderson A., Karathanasis S. K., Cate R., Zeitlin S., Chirgwin J. and Efstratiadis A. (1985) *Mol. Cell. Biol.*, **5**, 2090-2103.

Steiner D. F., Chan S. J., Welsh J. M. and Kwok S. C. M. (1985) *Ann. Rev. Genet.*, **19**, 463-484.

Stutz F. and Spohr G. (1987) *EMBO J.*, **6**, 1989-1995.

Stylianopoulou F., Efstratiadis A., Herbert J. and Pintar J. (1988) *Development*, **103**, 497-506.

Turner A. J. (1986) In *Essays in Biochemistry 22* (Ed. by Marchall, R. D. & Tipton, K. F.) pp.69-119, Academic Press, New York.

Yui R., Fujita T. and Ito S. (1980) *Biomed. Res.*, **1**, 42-46.

Vanin E. F. (1985) *Ann. Rev. Genet.*, **19**, 253-272.

Zapf J., Froesch E. R. and Humbel R. E. (1981) *Curr. Top. in Cell. Regul.*, **19**, 257-309.

Table 1: Evolutionary Distance per Nucleotide Site (K)*

| Comparison | | | | Chain or Peptide | | |
|---|---|---|---|---|---|---|
| | | | | A | B | C |
| bombyxin | A-1 | vs. | B-1 | 0.47 | 0.39 | 0.64 |
| | A-1 | vs. | C-1 | 0.40 | 0.75 | 0.54 |
| | B-1 | vs. | C-1 | 0.40 | 1.02 | 0.56 |
| human insulin | vs. rat insulin-I | | | 0.07 | 0.20 | 0.44 |

*K values were calculated using 3ST model (Kimura, 1981). Nucleotide sequences were referred from Iwami *et al.* (1989) for bombyxin A-1, Kawakami *et al.* (1989) for bombyxin B-1, Iwami *et al.* (1990) for bombyxin C-1, Bell *et al.* (1980) for human insulin, and Lomedico *et al.* (1979) for rat insulin-I. Reprinted from Iwami *et al.* (1990).

## FIGURE LEGENDS

**Figure 1.** Restriction map of a *Bombyx* genomic DNA segment in clone λ4K105 (Kawakami *et al.*, 1989). The closed and open boxes represent bombyxin family A (A-2 and A-3) and family B (B-1 and B-2) genes, respectively, and the arrows under the boxes show the direction of transcription. The stippled box represents a transposon-like inserted element (BIS). B, *Bam*HI; E, *Eco*RI; H, *Hin*dIII; K, *Kpn*I; P, *Pst*I; S, *Sal*I. Modified from Kawakami *et al.* (1989).

**Figure 2.** Schematic representation of the organization of genomic DNAs of bombyxin A-1 (Iwami *et al.*, 1989), A-2, A-3, B-1, B-2 (Kawakami *et al.*, 1989), C-1 (Iwami *et al.*, 1990), and human insulin (Bell *et al.*, 1980). Open boxes designate domains that are present in mature molecules and hatched boxes those absent. Numerals in boxes indicate the nucleotide number in the respective coding segments. The closed boxes represent TATA-boxes (TATA) and polyadenylylation signals [POLY(A)]. Human insulin gene contains two introns, one in the 5'-untranslated region and the other in the C-peptide (triangles). Paired or single basic amino acid residues by one-letter codes (KR, RR, R) indicate posttranslational proteolytic cleavage signals. ATG, transcription initiation codon; TAG, transcription termination codon; S, signal peptide; B, B-chain; C, C-peptide; A, A-chain.

**Figure 3.** Nucleotide sequences of the flanking regions of six bombyxin genes [A-1, from Iwami *et al.* (1989); A-2, A-3, B-1, and B-2, from Kawaka-

mi *et al.* (1989); C-1, from Iwami *et al.* (1990)]. Short direct repeats flanking the both sides of the presumed transcribed region are boxed, and the remnants of poly(A) stretch present in the family A genes are underlined. Reprinted from Iwami *et al.* (1990).

**Figure 4.** Amino acid sequence comparison between preprobombyxins (Iwami *et al.*, 1989; 1990; Kawakami *et al.*, 1989), bombyxins (Nagasawa *et al.*, 1984b; 1986; Maruyama *et al.*, 1988), human preproinsulin (Bell *et al.*, 1980), guinea-pig preproinsulin (Chan *et al.*, 1984), human preproIGF-I (Jansen *et al.*, 1983), preproIGF-II (Bell *et al.*, 1984), and human relaxin (Hudson *et al.*, 1983). Gaps are introduced for maximum alignment. Amino acids are numbered from the N-terminus of each domain of insulin. Boxes indicate the residues which are conserved among bombyxins except for preprobombyxin C-1. The residues which are completely conserved throughout the peptides shown and those conserved for all bombyxins are indicated by the open and solid circles, respectively. <Q, pyroglutamate.

**Figure 5.** Hydropathy profiles of preprobombyxin A-1, B-1, and C-1, and preproinsulin using the method of Kyte and Doolittle (1982). The values for every four sequential amino acids were calculated. S: signal peptide; B: B-chain; C: C-peptide; A: A-chain.

**Figure 6.** *in situ* hybridization of bombyxin mRNA.

A), A transverse section of a brain of 5th instar 0-day larva of *Bombyx mori* hybridized with bombyxin A-1 probe. Serial sections demonstrated that four

pairs of the dorso-medial neurosecretory cells of pars intercerebralis contained the bombyxin family-A mRNA. **B)**, A transverse section of a brain of 5th instar 0-day larva of *Bombyx* hybridized with bombyxin B-2 probe. Four pairs of the neurosecretory cells hybridized as in **A.** From Ikeno *et al.* (unpublished results).

**Figure 7.** Change in the amount of bombyxin mRNA in *Bombyx* brain during larval-pupal development. **A)**, **B)**, **C)**: Northern hybridization bands (0.6 kb, arrow) of total RNA (20 $\mu$g) from animals various days after 5th-instar ecdysis (V), wandering (W), and pupal ecdysis (P). Conditions for the hybridization was described in Adachi *et al.* (1989). Hybridization probes were gene A-1 for **A)**, gene B-1 for **B)**, and gene C-1 for **C)**. **D)**: Curves plotting the amounts of bombyxin A-, B-, and C-family mRNAs per brain for each developmental stage. The number of molecules was calculated as described (Adachi *et al.*, 1989). From Adachi *et al.* (unpublished results).

**Figure 8** Schematic drawing of a proposed evolutionary process of bombyxin genes. Solid boxes and wavy lines represent exons and introns, respectively. Proto-A, -B, and -C represent family-A, -B, and -C proto-bombyxin genes, respectively. To right is shown an alternative picture, for the processes distal to the step marked by asterisks, which proposes a different step where the processing of bombyxin genes occurred.

Fig. 1 IWAMI

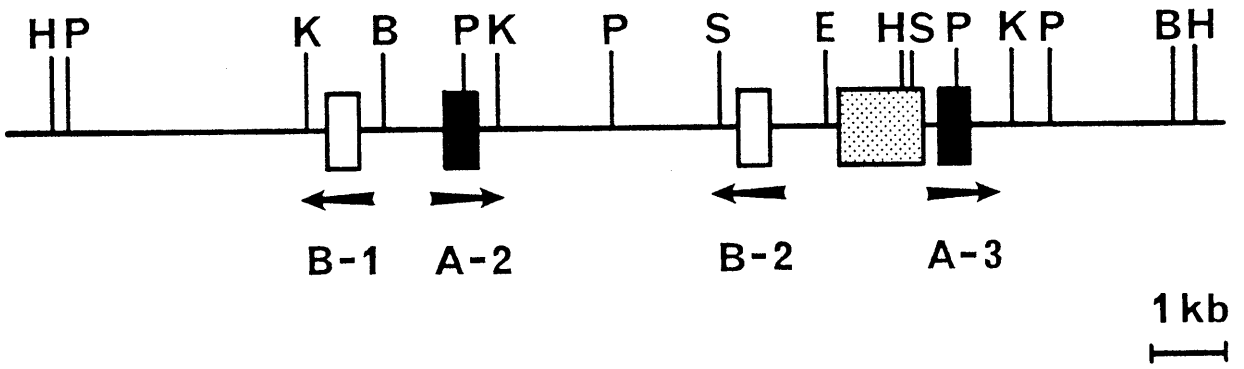


Fig. 2 IWAMI

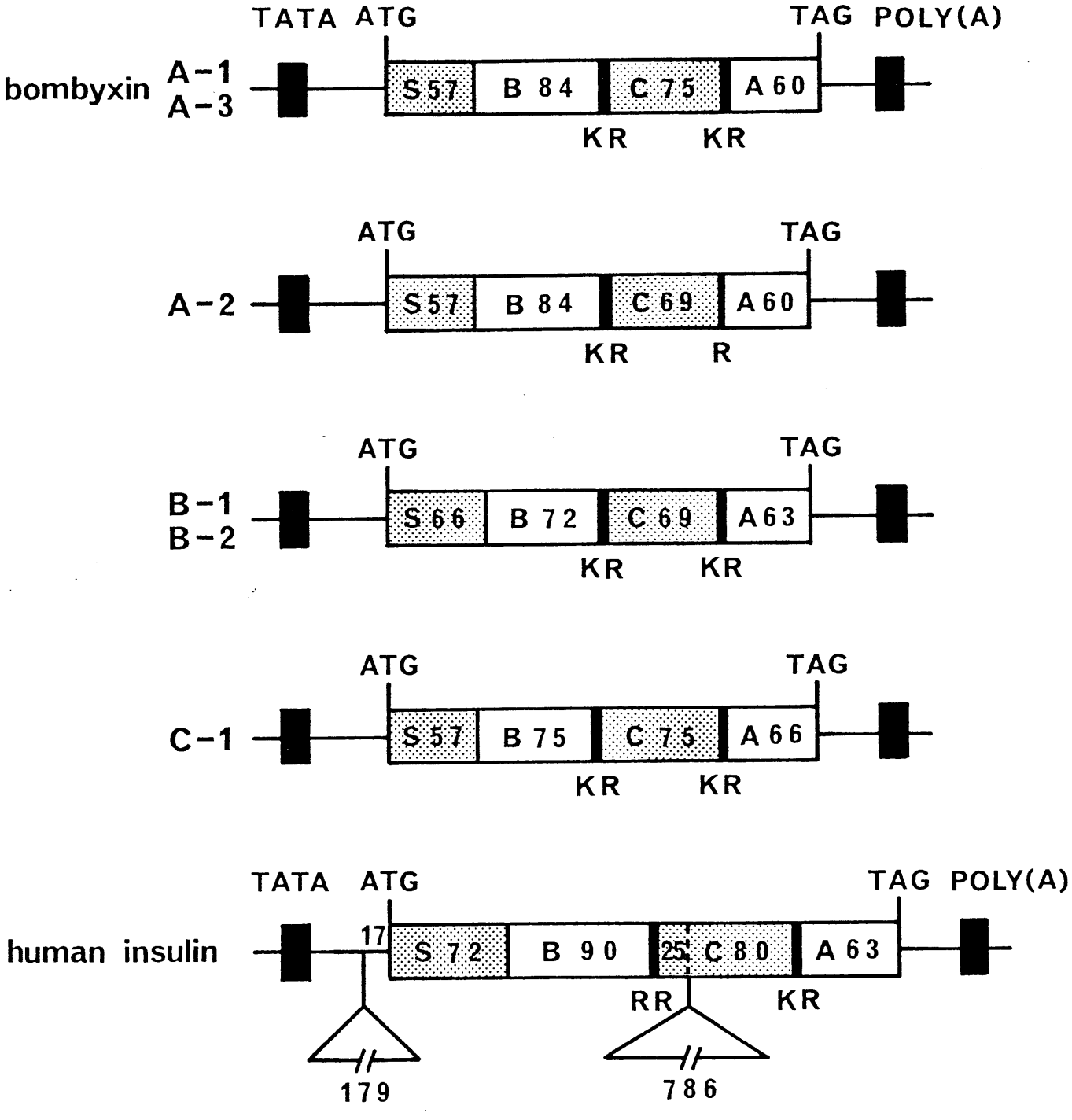

TATA
ATG
TAG
POLY(A)
bombyxin
A-1
A-3
S57
B 84
C 75
A 60
KR
KR
ATG
TAG
A-2
S57
B 84
C 69
A 60
KR
R
ATG
TAG
B-1
B-2
S 66
B 72
C 69
A63
KR
KR
ATG
TAG
C-1
S 57
B 75
C 75
A 66
KR
KR
TATA
ATG
TAG
POLY(A)
human insulin
17
S 72
B 90
25
C 80
A 63
RR
KR
179
786

Fig. 3 IWAMI

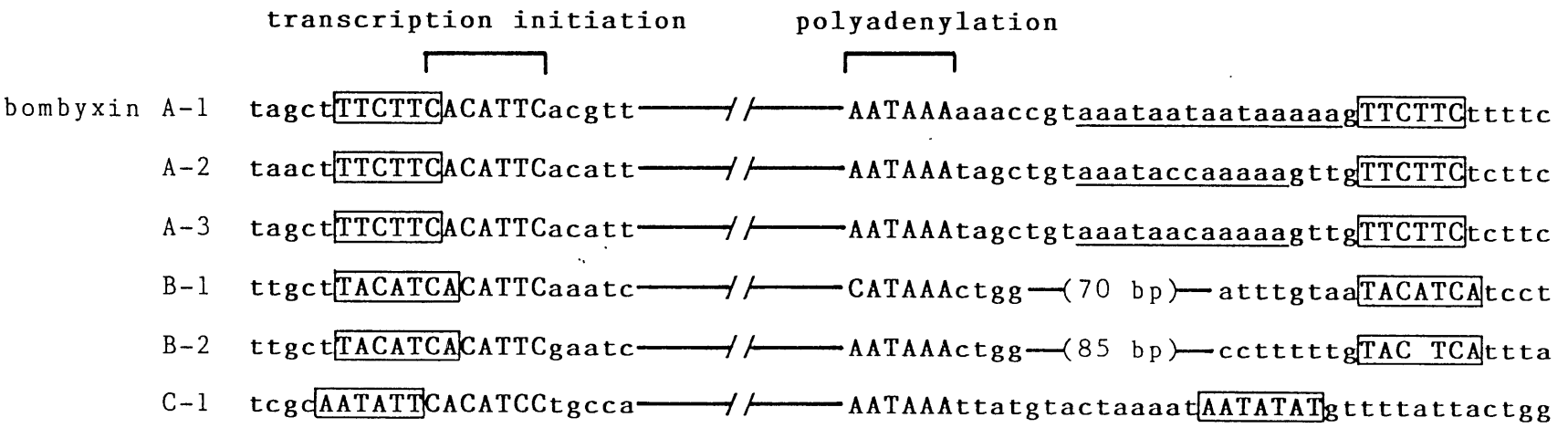

transcription initiation
polyadenylation
bombyxin A-1 tagctTTCTTCACATTCacgtt——//——AATAAAaaaccgtaaataataataaaaagTTCTTCttttc
A-2 taactTTCTTCACATTCacatt——//——AATAAAtagctgtaaataccaaaaagttgTTCTTCtcttc
A-3 tagctTTCTTCACATTCacatt——//——AATAAAtagctgtaaataacaaaaagttgTTCTTCtcttc
B-1 ttgctTACATCACATTCaaatc——//——CATAAActgg—(70 bp)—atttgtaaTACATCAtcct
B-2 ttgctTACATCACATTCgaatc——//——AATAAActgg—(85 bp)—ccttttgTAC TCAttta
C-1 tcgcAATATTCACATCCtgcca——//——AATAAAttatgtactaaaatAATATATgttttattactgg

Fig. 4 IWAMI

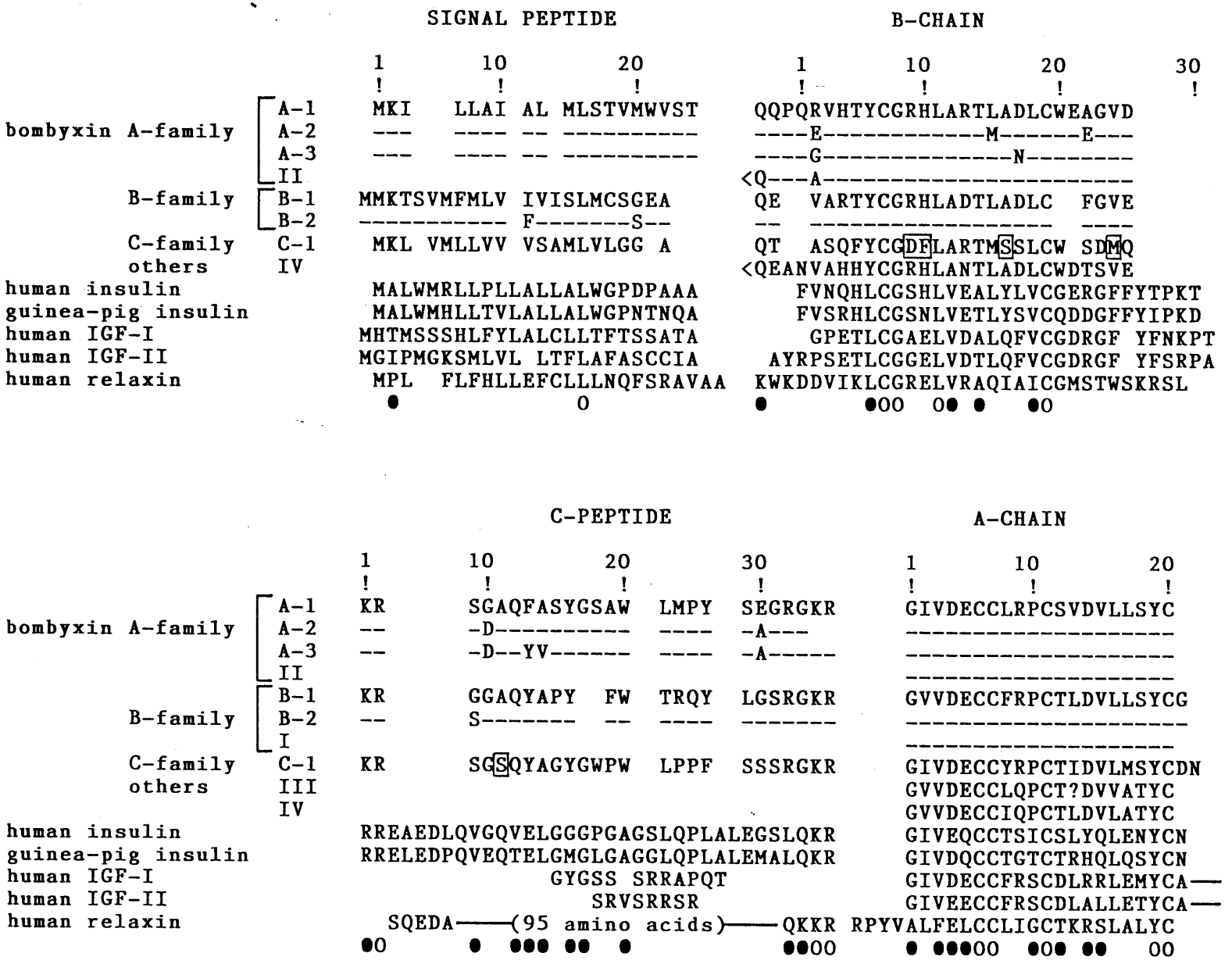

SIGNAL PEPTIDE
B-CHAIN
C-PEPTIDE
A-CHAIN
bombyxin A-family
A-1
A-2
A-3
II
B-family
B-1
B-2
I
C-family
C-1
others
III
IV
MKI LLAI AL MLSTVMWVST QQPQRVHTYCGRHLARTLADLCWEAGVD
MMKTSVMFMLV IVISLMCSGEA QE VARTYCGRHLADTLADLC FGVE
MKL VMLLVV VSAMLVLGG A QT ASQFYCGDFLARTMSSLCW SDMQ
MALWMRLLPLLALLALWGPDPAAA FVNQHLCGSHLVEALYLVCGERGFFYTPKT
MALWMHLLTVLALLALWGPNTNQA FVSRHLCGSNLVETLYSVCQDDGFFYIPKD
MHTMSSSHLFYLALCLLTFTSSATA GPETLCGAELVDALQFVCGDRGF YFNKPT
MGIPMGKSMLVL LTFLAFASCCIA AYRPSETLCGGELVDTLQFVCGDRGF YFSRPA
MPL FLFHLLEFCLLLNQFSRAVAA KWKDDVIKLCGRELVRAQIAICGMSTWSKRSL
KR SGAQFASYGSAW LMPY SEGRGKR GIVDECCLRPCSVDVLLSYC
KR GGAQYAPY FW TRQY LGSRGKR GVVDECCFRPCTLDVLLSYCG
KR SGSQYAGYGWPW LPPF SSSRGKR GIVDECCYRPCTIDVLMSYCDN
GVVDECCLQPCT?DVVATYC
GVVDECCIQPCTLDVLATYC
RREAEDLQVGQVELGGGPGAGSLQPLALEGSLQKR GIVEQCCTSICSLYQLENYCN
RRELEDPQVEQTELGMGLGAGGLQPLALEMALQKR GIVDQCCTGTCTRHQLQSYCN
GYGSS SRRAPQT GIVDECCFRSCDLRRLEMYCA
SRVSRRSR GIVEECCFRSCDLALLETYCA
SQEDA (95 amino acids) QKKR RPYVALFELCCLIGCTKRSLALYC

Fig. 5 IWAMI

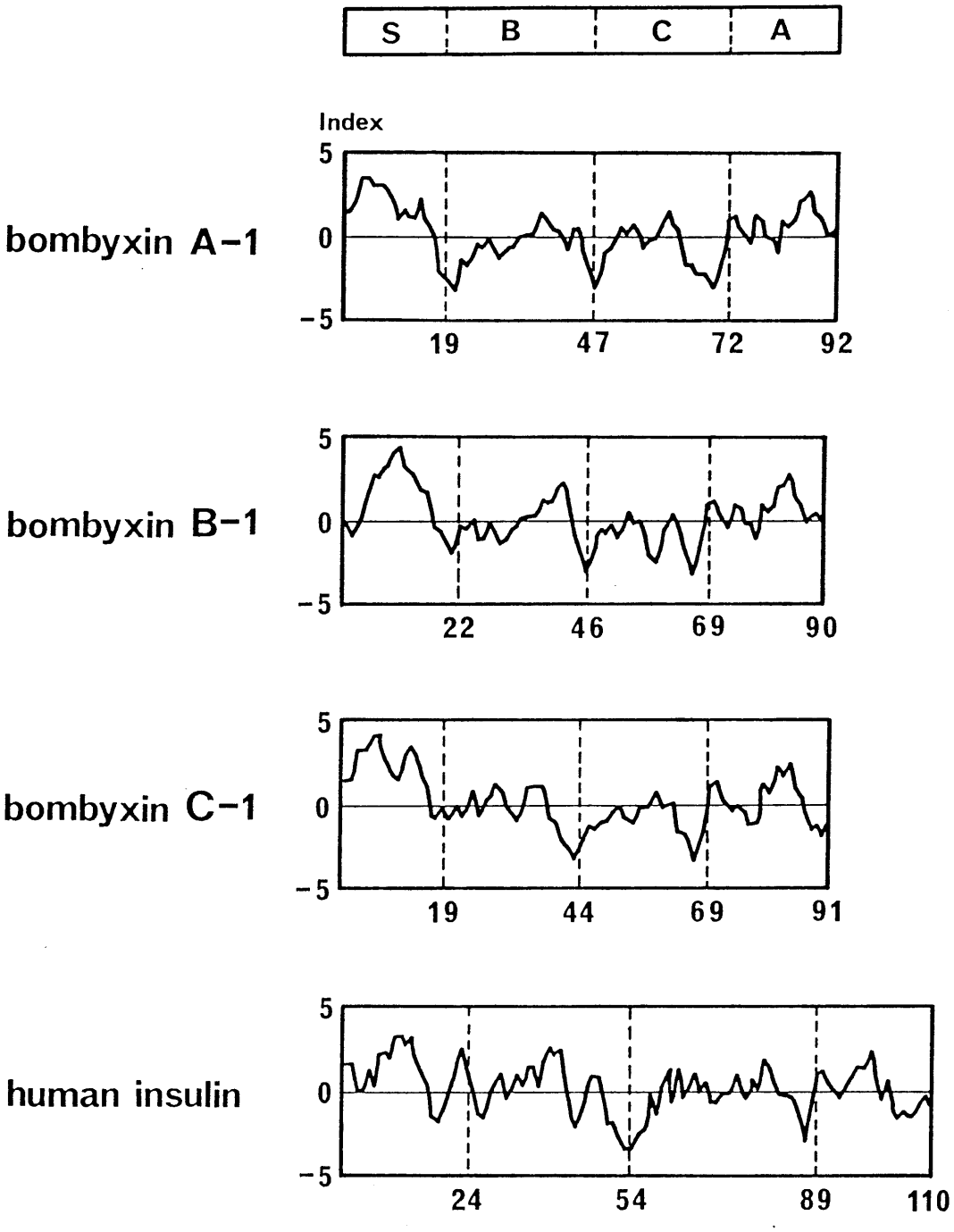
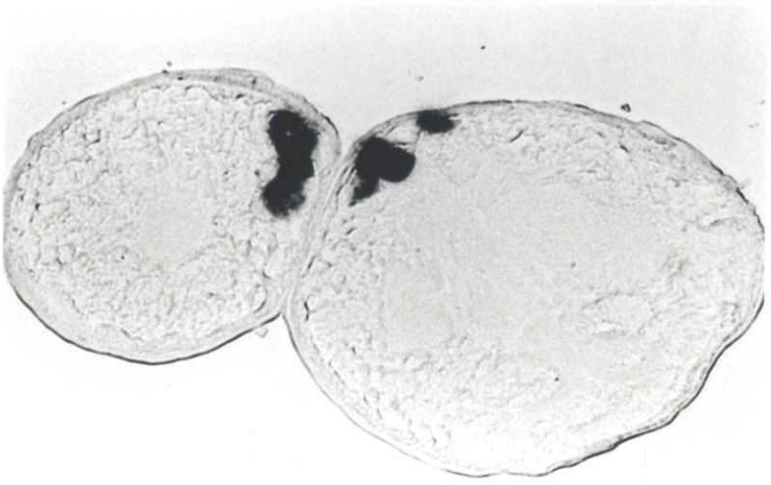

S
B
C
A
Index
bombyxin A-1
5
0
-5
19
47
72
92
bombyxin B-1
5
0
-5
22
46
69
90
bombyxin C-1
5
0
-5
19
44
69
91
human insulin
5
0
-5
24
54
89
110

Fig. 6

A)



B)

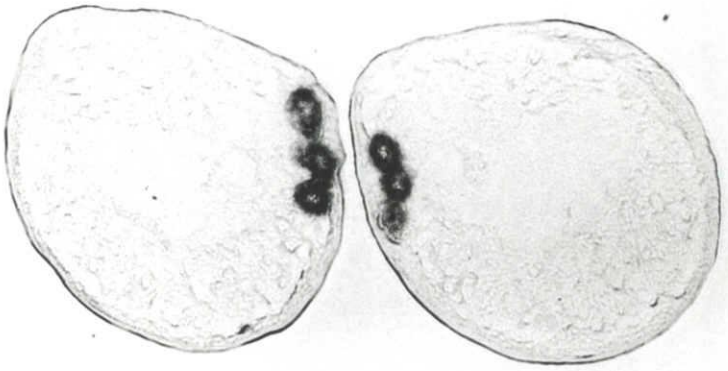


Fig. 7

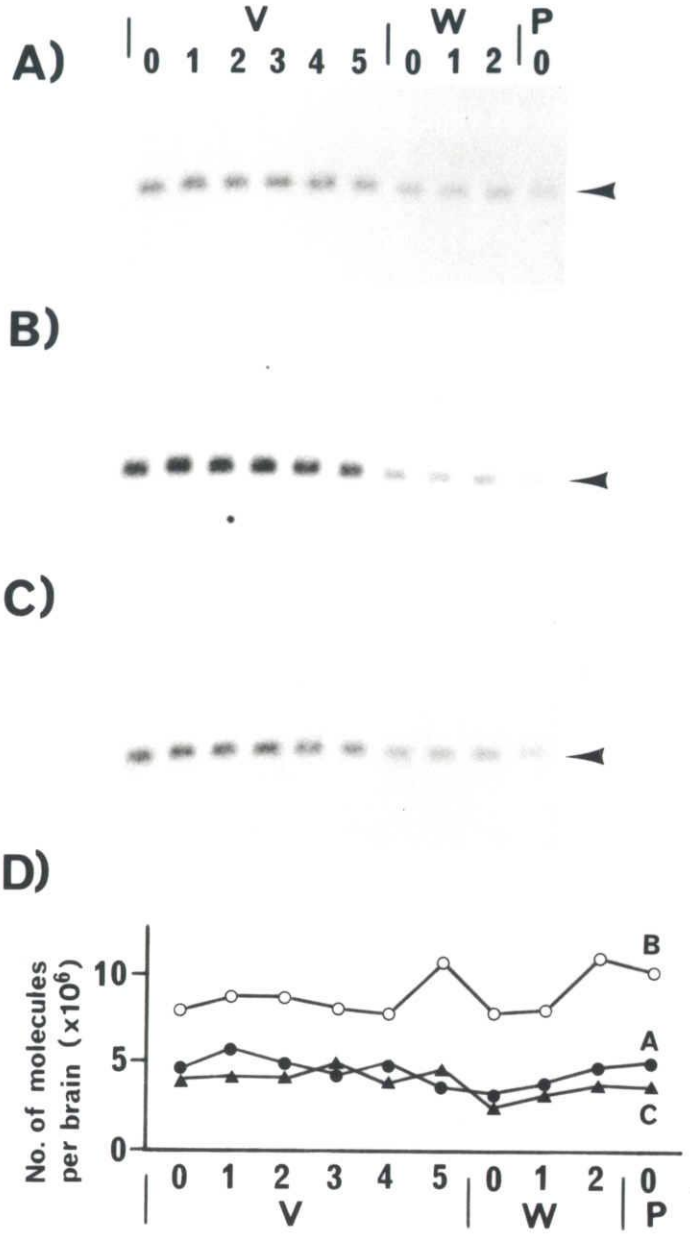
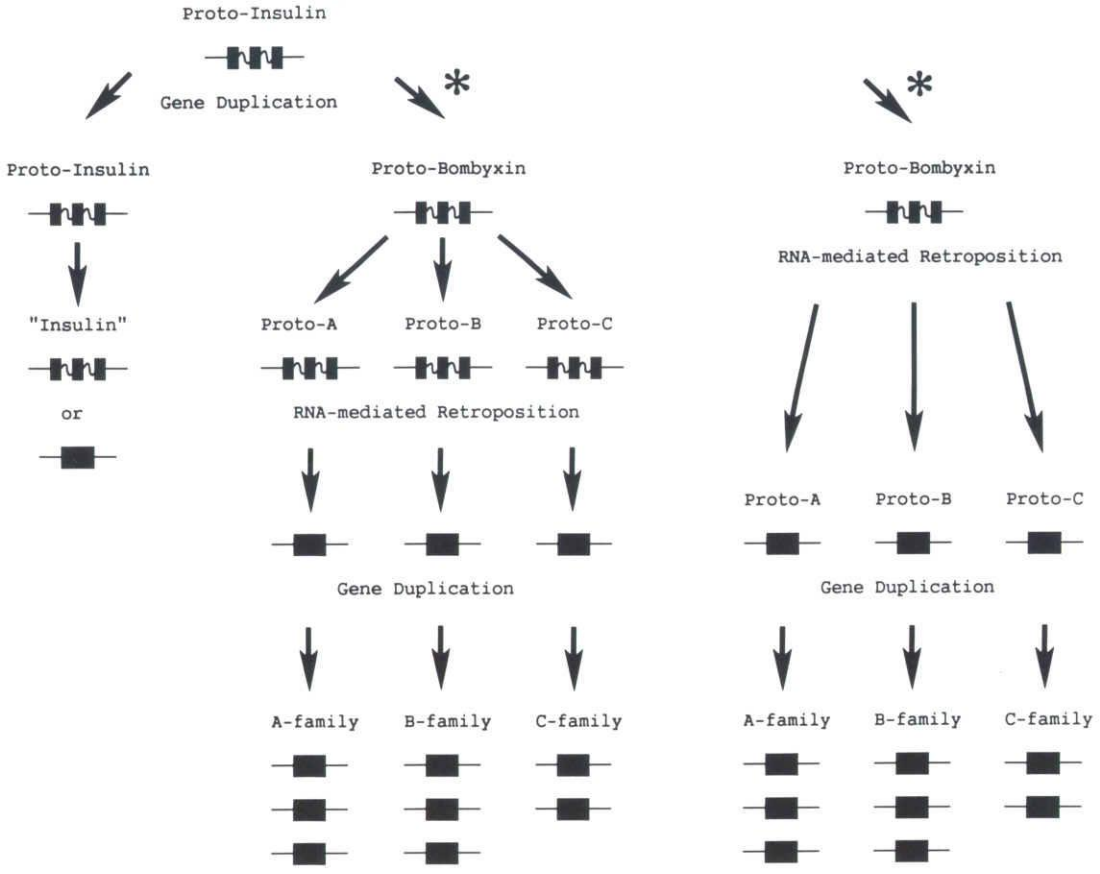

A)
V
W
P
0 1 2 3 4 5 0 1 2 0
B)
C)
D)
No. of molecules per brain (x10^6)
10
5
0
B
A
C
0 1 2 3 4 5 0 1 2 0
V
W
P

Fig. 8


Proto-Insulin
Gene Duplication
Proto-Insulin
"Insulin"
or
Proto-Bombyxin
Proto-A
Proto-B
Proto-C
RNA-mediated Retroposition
Gene Duplication
A-family
B-family
C-family
Proto-Bombyxin
RNA-mediated Retroposition
Proto-A
Proto-B
Proto-C
Gene Duplication
A-family
B-family
C-family

# 参考論文


1. Muto, A., Hori, H., Sawada, M., Kawauchi, Y., Iwami, M., Yamao, F. and Osawa, S., 1983. The ribosomal genes of Mycoplasma capricolum. Yale J. Biol. Med., 56, 373-376.

2. Sawada, M., Muto, A., Iwami, M., Yamao, F. and Osawa, S., 1984. Organization of ribosomal RNA genes in Mycoplasma capricolum. Mol. Gen. Genet., 196, 311-316.

3. Iwami, M., Muto, A., Yamao, F. and Osawa, S., 1984. Nucleotide sequence of the rrnB 16S ribosomal RNA gene from Mycoplasma capricolum. Mol. Gen. Genet., 196, 317-322.

4. Yamao, F., Muto, A., Kawauchi, Y., Iwami, M., Iwagami, S., Azumi, Y. and Osawa, S., 1985. UGA is read as tryptophan in Mycoplasma capricolum. Proc. Natl. Acad. Sci. U.S.A., 82, 2306-2309.

5. Andachi, Y., Yamao, F., Iwami, M., Muto, A. and Osawa, S., 1987. Occurrence of unmodified adenine and uracil at the first position of anticodon in threonine tRNAs in Mycoplasma capricolum. Proc. Natl. Acad. Sci. U.S.A., 84, 7398-7402.

6. Takeshita, T., Goto, Y., Nakamura, M., Fujii, M., Iwami, M., Hinuma, Y. and Sugamura, K., 1988. Phorbol esters can persistently replace interleukin-2 (IL-2) for the growth of a human IL-2-dependent T-cell line. J. Cell. Physiol., 136, 319-325.

7. Yasue, H., Iwami, M., Koide, Y., Ohtsubo, E. and Ishibashi, M., 1989. The oncogenicity of avian adenoviruses. IV. Confirmatory evidence for recombination between viral and cellular DNA sequences and repetition of the recombinant in cells of a tumor line. Virology, 169, 447-451.

8. Muto, A., Yamao, F., Sawada, M., Iwami, M., Hori, H. and Osawa, S. 1989. Ribosomal RNA genes in Mycoplasma capricolum. In Bacillus subtilis: Molecular biology and industrial application. (Eds. B. Maruo and H. Yoshikawa), pp.21-25. Kodansha, Tokyo, and Elsevier, Amsterdam.

9. Adachi, T., Takiya, S., Suzuki, Y., Iwami, M., Kawakami, A., Takahashi, S. Y., Ishizaki, H., Nagasawa, H. and Suzuki, A., 1989. cDNA structure and expression of bombyxin, an insulin-like brain secretory peptide of the silkmoth Bombyx mori. J. Biol. Chem., 264, 7681-7685.

10. Kawakami, A., Iwami, M., Nagasawa, H., Suzuki, A. and Ishizaki, H., 1989. Structure and organization of four clustered genes that encode bombyxin, an insulin-related brain secretory peptide of the silkmoth Bombyx mori. Proc. Natl. Acad. Sci. U.S.A., 86, in press.

11. 岩見雅史, 1989. 昆虫の変態をつかさどる遺伝子. 遺伝, 43. 60-66.

THE YALE JOURNAL OF BIOLOGY AND MEDICINE 56 (1983), 373-376

# The Ribosomal Genes of *Mycoplasma capricolum*

AKIRA MUTO, D.Sc., HIROSHI HORI, M.D., MAKOTO SAWADA, M.D.,*
YASUSHI KAWAUCHI, M.S.,* MASAFUMI IWAMI, B.S.,
FUMIAKI YAMAO, D.Sc., AND SYOZO OSAWA, D.Sc.

*Laboratory of Molecular Genetics, Department of Biology, Faculty of Science, Nagoya University, Nagoya, Japan*

Received January 4, 1983

The nucleotide sequence of 5S rRNA from *Mycoplasma capricolum* is more similar to that of the gram-positive bacteria than that of the gram-negative bacteria.

The presence of two copies of rRNA genes in *M. capricolum* genome has been demonstrated. The two different rRNA gene clusters have been cloned in *E. coli* plasmid vectors and analyzed for the rRNA gene organizations, demonstrating that the gene arrangement is in the order of 16S, 23S, and 5S rDNA.

The ribosomes of *M. capricolum* contain about 30 species of proteins in 50S and 20 in 30S subunits. The number and size of the ribosomal proteins are not significantly different from those of other eubacterial ribosomes.

## INTRODUCTION

The ribosome is an essential component in the cell, serving as a site of protein biosynthesis; thus all self-replicating cells have their own ribosomes. Many biochemical and genetic studies of ribosomal components have been done using various organisms, showing that the ribosomes exhibit divergent evolution. It is, therefore, advantageous to use ribosomes or ribosomal genes for studying evolution, since the results of one organism can be directly compared with those of others. The present paper summarizes our recent work on the ribosome of *Mycoplasma capricolum*. Several lines of evidence have suggested that the mycoplasma is phylogenetically related to gram-positive bacteria.

## MATERIALS AND METHODS

*M. capricolum* ATCC27343 (KID) was used throughout the experiments. The detailed experimental procedures were described in the separate papers [1-3].

## RESULTS AND DISCUSSIONS

### *The Nucleotide Sequence of 5S rRNA*

The sequences of 5S rRNAs from over two hundred organisms have been reported and used for deducing phylogenic relationships among them [4,5]. We have determined the total nucleotide sequence of *M. capricolum* 5S rRNA mainly by the rapid chemical degradation procedure of Peattie [6] and reported in a separate paper [1].



*On leave from Department of Biochemistry and Biophysics, Research Institute for Nuclear Medicine and Biology, Hiroshima University, Hiroshima, Japan.

Address reprint requests to: Dr. Akira Muto, Laboratory of Molecular Genetics, Dept. of Biology, Nagoya University, Furo-cho, Chikusa-ku, Nagoya, Japan 464

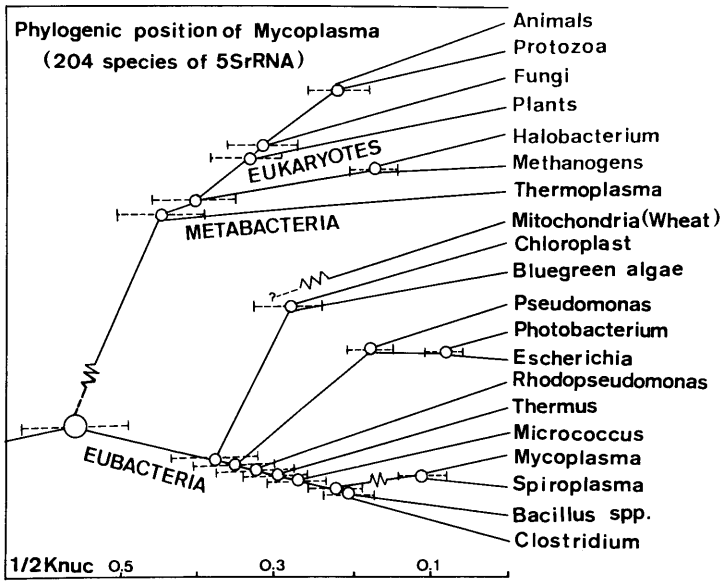



FIG. 1. Phylogenic tree of 5S rRNAs. **Knuc** represents the rate of nucleotide substitution (see [5]).

The length is 107 nucleotides, the shortest of all the 5S rRNAs so far known. The 5S rRNA has the lowest G + C content (42 percent among the 5S rRNAs yet sequenced, but still higher than the average G + C content of the total genome DNA (25 percent) of this species. The nucleotide sequence of *M. capricolum* 5S rRNA is more similar to that of the gram-positive bacteria than the gram-negative. Figure 1 shows the phylogenic relationship of *M. capricolum* to other organisms deduced by the sequence homologies of the 5S rRNAs (see [5]). Comparisons of the mycoplasma tRNA sequences with other bacterial tRNAs also revealed the same relationship [7]. Walker et al. [8] have sequenced 5S rRNA of *Spiroplasma* sp. BC3, indicating that the spiroplasma is closely related to the mycoplasma. On the other hand, *Thermoplasma acidophilum,* whose 5S rRNA sequence was determined by Luhrsen et al. [9], has been shown to be a member of Metabacteria [5].

*The Organization of rRNA Genes*

It has been previously shown that *M. capricolum* has only one set of rRNA genes by hybridization-saturation experiments between 16S and 23S rRNAs, and DNA [10]. We reexamined the copy number of rRNA genes by the Southern blotting analysis [11]. The total DNA was completely digested either with *Eco*RI, *Bg*lII, or *Xba*I endonucleases, separated by agarose gel electrophoresis, and submitted to hybridization with $^{32}$P-3′-end-labeled 16S, 23S, and 5S rRNAs, respectively. The results have clearly shown that the *M. capricolum* genome carries at least two sets of genes for 16S, 23S, and 5S rRNAs [2]. For example, the *Bg*lII digested DNA gel gave two distinct bands of 6.8 and 9.8 kilobases (Kb) long, equally with 16S, 23S, and 5S rRNA. The results also suggest that each one of the copies for 16S, 23S, and 5S rRNA genes are clustered on the chromosome. To see the rRNA gene organization, we have cloned several DNA fragments containing *M. capricolum* rRNA genes to the *Escherichia coli* plasmids pBR322 or pBR325. The gene organization in the cloned DNA fragments was analyzed by the Southern blotting method. Figure 2 shows the physical maps of two of these hybrid plasmid DNAs. The plasmids pMCB221 and pMCB339, respectively, contain 6.8 and 9.8 Kb DNA fragments

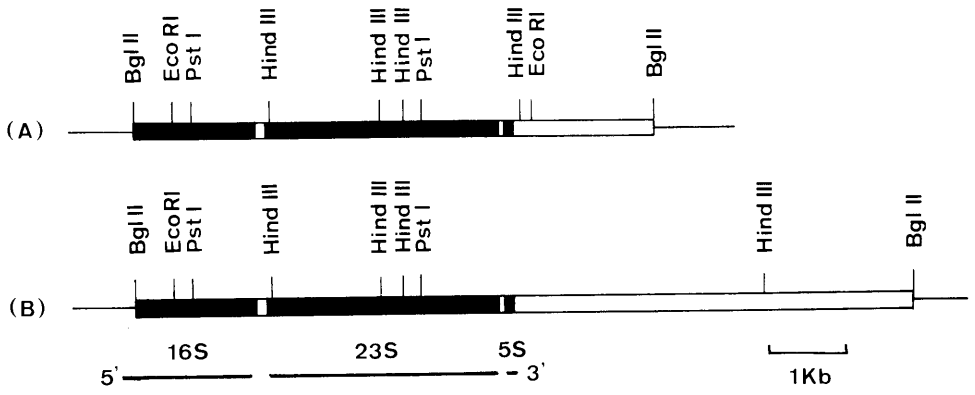



FIG. 2. Physical maps of *M. capricolum* rRNA genes. The DNA fragments generated by *Bg*lII digestion of *M. capricolum* genome were cloned in the plasmid vector pBR322. The hybrid plasmids (**A**) pMCB221 and (**B**) pMCB339, respectively, contain 6.8 and 9.8 Kb DNA fragments including rRNA genes. Thick lines represent the *M. capricolum* DNA. The vector DNA is shown as thin lines. The dark zones represent rRNA-coding regions.

generated by *Bg*lII digestion, each including a different set of rRNA genes. In both DNAs, the rRNA genes are arranged in the order of 16S, 23S, and 5S rDNA. Thus the arrangement of *M. capricolum* rRNA genes is common to other prokaryotic rRNA gene clusters so far known. We have determined the nucleotide sequences of several parts of the cloned rRNA genes, including the spacer region between 16S and 23S rRNA genes. The results have revealed that the coding sequences of 16S and 23S rRNA genes of *M. capricolum* are highly similar to those of *Bacillus subtilis* and *E. coli,* suggesting that the rRNA genes are well conserved in these prokaryote genomes. On the other hand, the spacer sequences between 16S and 23S rRNA genes have little resemblence between *M. capricolum* and *B. subtilis* or *E. coli.* We have so far not detected any tRNA like-sequence in the spacer region.

The isolation of the two different clones, each containing a set of rRNA genes, strongly supports the idea that the *M. capricolum* genome carries two sets of rRNA genes [2]. Contrary to this, Amikam et al. [12] recently reported that *M. capricolum* contains only one set of rRNA genes, while there are two sets in *M. mycoides* subsp. *capri* and *Acholeplasma laidlawii.* The reason for this discrepancy is not clear. The genome size of mycoplasmas is one-quarter to one-fifth of that of *E. coli* or *B. subtilis.* It is known that the rRNA gene copies are at least seven in *E. coli* [13] and ten in *B. subtilis* [14]. These facts suggest that the copy number of rRNA genes varies from one bacterial species to another roughly in proportion to their genome size.

*Ribosomal Proteins*

The total and ribosomal proteins of *M. capricolum* were analyzed by two-dimensional (2D) gel electrophoresis, and compared with those of *E. coli* and *B. subtilis* [3]. First, the total proteins were separated by O'Farrell's 2D gel system [15,16] to estimate the approximate number of proteins in the cell. The number of protein spots detected was about 350, indicating that the *M. capricolum* genome contains at least about 350 genes for proteins. The number of whole cell protein spots of *E. coli* or *B. subtilis* detected under the same conditions is about 1,100.

Second, the ribosomal proteins were analyzed by the 2D gel system of Kaltschmidt and Wittmann [17]. The number of the *M. capricolum* ribosomal protein species is

at least 30 for 50S and 20 for 30S subunits with average molecular weight of about 15,000 daltons. This shows that the number and size of *M. capricolum* ribosomal proteins are not significantly different from other eubacterial ribosomal proteins in contrast to a great reduction of the number of total proteins. This suggests that the number and size of the genes for ribosomal proteins are conserved in the mycoplasma genome in spite of the limited genetic capacity. The protein profiles of both 30S and 50S subunits in 2D gel electrophoresis, as a whole, resemble those of *B. subtilis* or *B. stearothermophilus* rather than *E. coli.* Several characteristic features of ribosomal proteins from gram-positive bacteria, for example, the lack of protein S1 (see [18,19]), are also seen in *M. capricolum.* These observations are consistent with the idea that the mycoplasma is phylogenetically close to *Bacilli.*

ACKNOWLEDGEMENTS

This work was supported by the Naito Science Foundation Research Grant for 1981 and grants from the Ministry of Education of Japan (No. 56480377 and No. 57121003).

## REFERENCES

1. Hori H, Sawada M, Osawa S, et al: The nucleotide sequence of 5S rRNA from *Mycoplasma capricolum.* Nucl Acids Res 9: 5407–5410, 1981
2. Sawada M, Osawa S, Kobayashi H, et al: The number of ribosomal RNA genes in *Mycoplasma capricolum.* Molec Gen Genet 182:502–504, 1981
3. Kawauchi Y, Muto A, Osawa S: The protein composition of *Mycoplasma capricolum.* Molec Gen Genet 188:7–11, 1982
4. Hori H, Osawa S: Evolutionary change in 5S RNA secondary structure and a phylogenic tree of 54 5S RNA species. Proc Natl Acad Sci USA 76:381–385, 1979
5. Hori H, Itoh T, Osawa S: The phylogenic structure of the Metabacteria. Zbl Bakt Hyg, I Abt Orig C3:18–30, 1982
6. Peattie DA: Direct chemical method for sequencing RNA. Proc Natl Acad Sci USA 76:1760–1764, 1979
7. Kilpatrick MW, Walker RT: The nucleotide sequence of glycin tRNA from *Mycoplasma mycoides* sp. *capri.* Nucl Acids Res 8:2783–2786, 1980
8. Walker RT, Chelton ETJ, Kilpatrick MW, et al: The nucleotide sequence of the 5S rRNA from *Spiroplasma* species BC3 and *Mycoplasma mycoides* sp. *capri* PG3. Nucl Acids Res 10:6363–6367, 1982
9. Luehrsen KR, Fox GE, Kilpatrick MW, et al: The nucleotide sequence of the 5S RNA from the archaebacterium *Thermoplasma acidophilum.* Nucl Acids Res 9:965–970, 1981
10. Ryan JL, Morowitz HJ: Partial purification of native rRNA and tRNA cistrons from *Mycoplasma* sp. (KID). Proc Natl Acad Sci USA 63:1282–1289, 1969
11. Southern EM: Detection of specific sequence among DNA fragments separated by gel electrophoresis. J Mol Biol 98:503–517, 1975
12. Amikam D, Razin S, Glaser G: Ribosomal RNA genes in Mycoplasma. Nucl Acids Res 10:4215–4222, 1982
13. Kenerly ME, Morgan EA, Post L, et al: Characterization of hybrid plasmid carrying individual ribosomal ribonucleic acid transcription units of *Escherichia coli.* J Bacteriol 132:931–949, 1977
14. Kobayashi H, Osawa S: The number of 5S rRNA genes in *Bacillus subtilis.* FEBS Lett 141:161–163, 1982
15. O'Farrell PH: High resolution two-dimensional electrophoresis of proteins. J Biol Chem 250:4007–4021, 1975
16. O'Farrell PZ, Goodmann HM, O'Farrell PH: High resolution two-dimensional electrophoresis of basic as well as acidic proteins. Cell 12:1133–1142, 1977
17. Kaltschmidt E, Wittmann HG: Ribosomal proteins VII. Two dimensional polyacrylamide gel electrophoresis for fingerprinting of ribosomal proteins. Anal Biochem 36:402–412, 1970
18. Isono K, Isono S: Lack of ribosomal protein S1 in *Bacillus stearothermophilus.* Proc Natl Acad Sci USA 73:767–770, 1976
19. Higo K, Otaka E, Osawa S: Purification and characterization of 30S ribosomal proteins from *Bacillus subtilis:* Correlation to *Escherichia coli* 30S proteins. Molec Gen Genet 185:239–244, 1982

Mol Gen Genet (1984) 196:311–316

MGG



# Organization of ribosomal RNA genes in *Mycoplasma capricolum*

**Makoto Sawada*, Akira Muto, Masafumi Iwami, Fumiaki Yamao, and Syozo Osawa**

Laboratory of Molecular Genetics, Department of Biology, Faculty of Science, Nagoya University, Furo-Cho, Chikusa-Ku, Nagoya 464, Japan

**Summary.** DNA segments carrying rRNA genes of *Mycoplasma capricolum* have been cloned and characterized by restriction endonuclease mapping, DNA-RNA hybridization and nucleotide sequencing. The *M. capricolum* genome has two sets of rRNA gene clusters, where the arrangement is in the order of (5′)16S-23S-5S(3′). The spacer region between 16S and 23S rDNA is extremely rich in AT and does not carry any tRNA genes.

## Introduction

Mycoplasmas are parasitic prokaryotes and the smallest free-living organisms known (for reviews, see Maniloff and Morowitz 1972; Razin 1978). Their genome size is about $0.5 \times 10^9$ daltons, and is about one-fifth of that of *Escherichia coli*. We have estimated the number of the genes in the mycoplasma genome to be only about 400 (Kawauchi et al. 1982). Furthermore, the GC contents of DNA from most of the mycoplasma species are extremely low as compared with those of other bacterial DNAs. In spite of these characteristic features, little is known concerning the the structure, organization and expression of the mycoplasma genes. This is mainly due to the difficulty of obtaining available genetic markers, and to the lack of a genetic recombination system in the organisms (see Stanbridge and Reff 1979). However, the recent progess in gene cloning and DNA sequencing techniques has provided useful tools for studying the genetics of mycoplasmas. As a first step in genetic analyses of mycoplasmas, we have attempted to elucidate the organization of rRNA genes in *Mycoplasma capricolum*. In a previous report, we demonstrated that the *M. capricolum* genome contains at least two sets of genes for 16S, 23S and 5S rRNAs (Sawada et al. 1981). In the present paper, we have constructed a library of genomic DNA segments from *M. capricolum* and identified the recombinant plasmids carrying cloned rRNA gene clusters. The gene organization in each cluster has been analyzed and compared with those of other prokaryotic rRNA genes.

* On leave from Department of Biochemistry and Biophysics, Research Institute for Nuclear Medicine and Biology, Hiroshima University, Kasumi-Cho, Minami-Ku, Hiroshima 734, Japan. *Present address:* Division of Hematology and Immunology of Internal Medicine, Kanazawa Medical University, Uchinada-Cho, Kahoku-Gun, Ishikawa Pref. 920-02, Japan

*Offprint request to:* A. Muto

## Materials and methods

*a) Materials.* The restriction endonuclease *Mbo*II was purchased from Bethesda Research Laboratories, Inc. (Maryland, USA). Other restriction endonucleases, bacterial alkaline phosphatase, $T_4$-DNA ligase and $T_4$-polynucleotide kinase were purchased from Takara Shuzo Co. Ltd. (Kyoto, Japan). Proteinase K and ribonuclease A were from Sigma Chemical Company (St. Louis, USA). Phosphorus-32 was from Amersham Japan (Tokyo, Japan). Nitrocellulose filters were purchased from Sartorius GmbH (Gottingen, FRG) and Millipore Corporation (Massachusetts, USA).

*b) Preparation of DNA and rRNA. Mycoplasma capricolum* ATCC 27343 was grown as described previously (Sawada et al. 1981; Kawauchi et al. 1982).

For the preparation of high molecular weight DNA, cells were harvested by centrifugation, suspended in TE-buffer (10 mM Tris-HCl, 1 mM EDTA, pH 7.6) and lysed at 37° C for 2 h in the presence of 0.5% SDS and 100 μg/ml proteinase K. The lysate was treated twice with TE-buffer-saturated phenol. The aqueous phase was dialyzed against TE-buffer overnight and treated with 50 μg/ml ribonuclease A at room temperature for 1 h. DNA was then purified by the proteinase K treatment, as described above.

Total rRNA was isolated from 70S ribosomes, and 16S, 23S, and 5S rRNAs were prepared from 30S and 50S ribosomal subunits as described in Sawada et al. (1981). Ribosomes and their subunits were isolated by sucrose gradient centrifugation as described in Kawauchi et al. (1982). 5S rRNA was further purified by polyacrylamide-7 M urea gel electrophoresis. The purified RNA preparations were hydrolyzed to give an average chain length of about 100 nucleotides and treated by bacterial alkaline phosphatase. The 5′-ends were then labeled with [$\gamma$-$^{32}$P] ATP using polynucleotide kinase according to Maxam and Gilbert (1977). Radioactive ATP was prepared according to Walseth and Johnson (1979).

*c) DNA-cloning. M. capricolum* DNA was completely digested with either *Eco*RI, *Hin*dIII or *Bgl*II under the conditions recommended by the manufacturers. The *Eco*RI-digested DNA fragments were ligated with *Eco*RI-digested and phosphatase-treated vector pBR325. Similarly, the *Hin*dIII- or *Bgl*II-digested fragments were ligated with *Hin*dIII- or *Bam*HI-digested pBR322. The ligated DNAs were used for transformation of *E. coli* HB101. Ampicillin-resistant

and chloramphenicol- (for pBR325) or tetracycline- (for pBR322) sensitive clones were picked up, purified and used for colony hybridization analyses using 5′-$^{32}$P-labeled total rRNA as a probe (Grunstein and Hogness 1975). Positive clones were subjected to a second colony hybridization for confirmation.

For the sequencing of the spacer regions between 16S and 23S rRNA genes (see Section *c* of Results), subclones containing the spacer regions were constructed from an *Eco*RI/*Hin*dIII DNA fragment of 1.2 kilo base pairs (kb) using the *Eco*RI/*Hin*dIII double-digested pBR322 as a vector (see Results).

*d) Physical mapping.* Plasmid DNAs were isolated from the transformed cells according to Oka et al. (1981). The DNA was digested with endonucleases and separated by 0.7% or 1% agarose gel electrophoresis using an E-buffer system (40 mM Tris-acetate, 20 mM acetate, 2 mM EDTA, pH 7.9). *Eco*RI and/or *Hin*dIII digests of lambda phage DNA ($\lambda_{cI857S7}$) were used as size markers. Gels were photographed under 365nm ultraviolet light using a red filter and the separated fragments were denatured, neutralized and transferred to a nitrocellulose filter (Sartorius SM 71) according to Southern (1975). The DNA fragments on the filters were hybridized with 5′-$^{32}$P-labeled 16S, 23S or 5S rRNA as described (Sawada et al. 1981).

*e) Sequencing of DNA fragments.* DNA fragments separated by electrophoresis were eluted from the gels by the glass beads method (Vogelstein and Gillespie 1979). Nucleotide sequences were determined by the method of Maxam and Gilbert (1977).

## Results

*a) Cloning of rRNA genes*

A library of the *M. capricolum* genomic clones was constructed by insertion of *Bgl*II-, *Eco*RI- or *Hin*dIII-digested fragments of the DNA into plasmid pBR322 or pBRR325 using *E. coli* HB101 as a host. A total of 2510 clones (610 clones in the *Bgl*II, 1235 in the *Eco*RI and 665 in the *Hin*dIII libraries) was isolated. All clones were examined by colony hybridization for the presence of the rRNA sequences. About 40 positive clones were detected. The plasmids were isolated and analyzed for the structures by restriction endonuclease digestions. Figure 1(a) shows restriction maps of the representative recombinant plasmids carrying rRNA genes. From the *Bgl*II library, two kinds of plasmid containing respectively a 6.8 kb (pMCB230) and a 9.8 kb (pMCB1002) insert were isolated. The Southern hybridization analysis of the *Bgl*II-digested genomic DNA gave two distinct bands of 6.8 kb and 9.8 kb that hybridized with the 3′-terminally labeled 16S, 23S and 5S rRNAs (Sawada et al. 1981). It is thus clear that the inserts in the pMCB230 and pMCB1002 are different, each containing a set of rRNA genes originating from different parts of the chromosome. The rRNA gene clusters included in the 6.8 kb and 9.8 kb fragments are tentatively designated *rrnA* and *rrnB*, respectively.

From the *Eco*RI library, a plasmid containing a 4.5 kb (pMCE223) *Eco*RI fragment was isolated. The 4.5 kb insert in pMCE223 is a part of the 6.8 kb insert in pMCB230 (*rrnA*), because the *Eco*RI digestion of the pMCB230 generated a 4.5 kb DNA segment that revealed a restriction map identical to the insert in pMCE223. On the other hand,

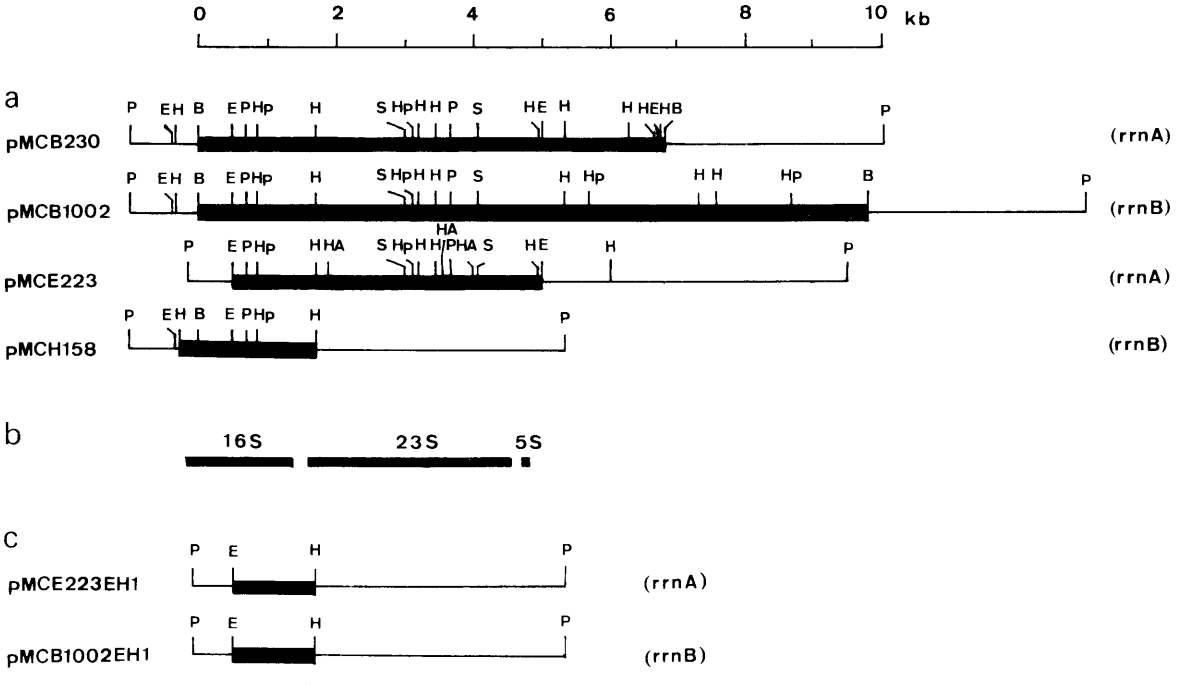



**Fig. 1.** (a) Restriction maps of the pMCB230, pMCB1002, pMCE223 and pMCH158. Thick lines represent the inserted *M. capricolum* DNA fragments and thin lines show vectors, pBR322 or pBR325. "B" represents *Bam*HI site (pBR322) or *Bgl*II site (*M. capricolum*). E: *Eco*RI, H: *Hin*dIII, HA: *Hae*III, Hp: *Hpa*I, p: *Pst*I and S: *Sma*I. *Hae*III cutting sites were determined only in pMCE223 4.5 kb *Eco*RI fragment. (b) The locations of rRNA genes on the cloned DNA segments. Locations of the 3′-end of 23S rRNA genes and the 5S rRNA gene are approximate. (c) The subcloned DNA segments in pMCB1002EH1 and pMCE223EH1 (see Fig. 3)

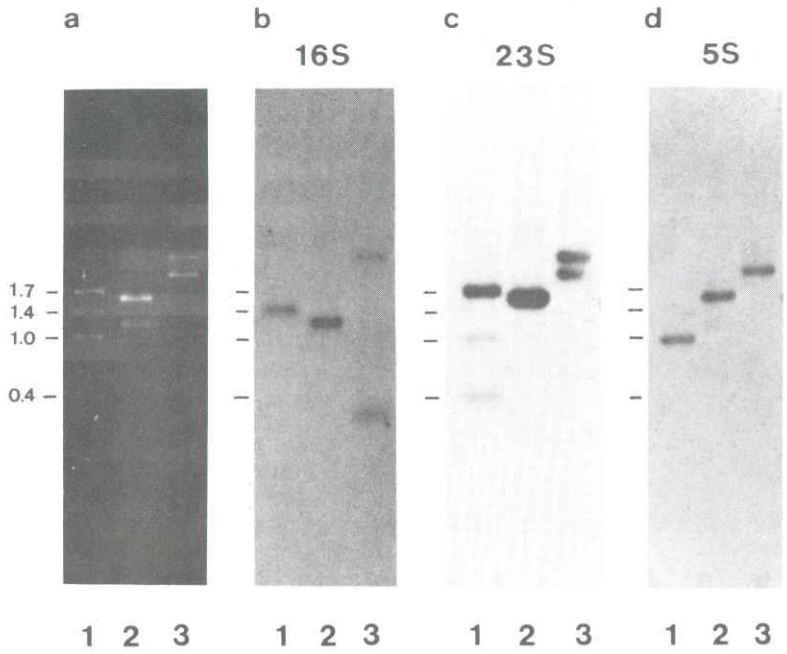



**Fig. 2a–d.** Southern hybridization of a 4.5 kb fragment of pMCE223 with $^{32}$P-labeled rRNAs. The 4.5 kb *Eco*RI insert of pMCE223 was isolated and digested by *Hae*III (lane 1), *Hin*dIII (lane 2) and *Hpa*I (lane 3), respectively. The products were separated by 1% agarose gel electrophoresis, stained by ethidium bromide **a**, and hybridized with $^{32}$P-labeled 16S **b**, 23S **c** and 5S **d** rRNAs, respectively. The sizes (kb) of *Hae*III digestion products are shown in the left side of **a**

the 9.8 kb *Bgl*II insert in pMCB1002 (*rrnB*) had only one *Eco*RI cutting site (see Fig. 1a). Genomic Southern hybridization also showed the presence of a 4.5 kb *Eco*RI fragment that hybridized with 16S, 23S, and 5S rRNAs (Sawada et al. 1981).

Plasmid pMCH158 having a 2.0 kb DNA insert was isolated from the *Hin*dIII library. The restriction maps showed that the 2.0 kb insert overlapped with the insert in pMCE223, pMCB230 or pMCB1002, as shown in Fig. 1(a).

*b) Physical mapping*

The 4.5 kb *Eco*RI fragment was isolated from pMCE223 (*rrnA*) and digested separately with *Hae*III, *Hin*dIII and *Hpa*I. The products were separated by 1% agarose gel electrophoresis, transferred to nitrocellulose filters and hybridized with $^{32}$P-labeled 16S, 23S or 5S rRNA (Fig. 2). The results are summarized in Table 1. An *Hae*III digestion of the 4.5 kb *Eco*RI fragment generated four fragments of 1.7, 1.4, 1.0, and 0.4 kb. The 1.4 kb fragment hybridized with both 16S and 23S rRNAs and the 1.0 kb fragment with both 23S and 5S rRNAs, while the 1.7 and 0.4 kb fragments hybridized only with 23S rRNA (Fig. 2, lane 1). Similarly the *Hap*II digestion gave a 2.2 kb fragment hybridizable with both 16S and 23S rRNA and a 1.9 kb fragment with 23S and 5S rRNAs (Fig. 2, lane 3). These results show the physical linkages between 16S and 23S rRNA genes and between 23S and 5S rRNA genes. The hybridization of rRNAs with *Hin*dIII digestion products also supports this conclusion (Fig. 2, lane 2). These results and the locations of the hybridized fragments on the restriction map of the 4.5 kb *Eco*RI fragment (Fig. 1a) indicate that the rRNA genes are arranged in the order of 16S-23S-5S as shown in Fig. 1 (b). The location of 5′-ends of the 16S and 23S rRNA genes and the 3′-end of the 16S rRNA gene was estimated by nucleotide sequencings (see section *c* and Iwami et al. 1984). Although the precise location of the 5S rRNA gene has not yet been determined, a preliminary nucleotide sequencing revealed that the sequence homologous to the *M. capricolum* 5S rRNA (Hori et al. 1981) is located at about the region 3 kb downstream from the 5′-end of the 23S rRNA gene on the same strand (unpublished result).

**Table 1.** Southern hybridization of a 4.5 kb fragment of pMCE222 with $^{32}$P-labeled rRNAs

| Enzyme | Products (kb) | Hybridization[a] | | |
|---|---|---|---|---|
| | | 16S | 23S | 5S |
| *Hae*III | 1.7 | − | + | − |
| | 1.4 | + | (+) | − |
| | 1.0 | − | + | + |
| | 0.4 | − | + | − |
| *Hin*dIII | 1.55[b] | − | + | + |
| | 1.5[b] | − | + | − |
| | 1.2 | + | (+) | − |
| | 0.3 | − | (+) | − |
| *Hpa*I | 2.2 | + | + | − |
| | 1.9 | − | + | + |
| | 0.4 | + | − | − |

[a] +: strongly hybridized; (+): faintly hybridized
[b] These two fragments could not be resolved in the gels

Southern hybridization experiments were also carried out with plasmid pMCB1002 DNA containign *rrnB* (Fig. 3a–d). The results indicated that the rRNA gene arrangement in *rrnB* was the same as that in *rrnA* (Fig. 1b). Furthermore, the restriction maps of the estimated coding regions of *rrnA* and *rrnB* resembled each other (Fig. 1a).

The 2.0 kb *Hin*dIII insert in pMCH158 hybridized strongly with 16S and weakly with 23S rRNAs (data not shown). The nucleotide sequencing analysis of this fragment demonstrated that the insert included the following parts of *rrn*B: the 5′-flanking sequence of the 16S rRNA gene, a full length of the 16S rRNA gene and the 5′-terminal sequence of the 23S rRNA gene (see Iwami et al. 1984).

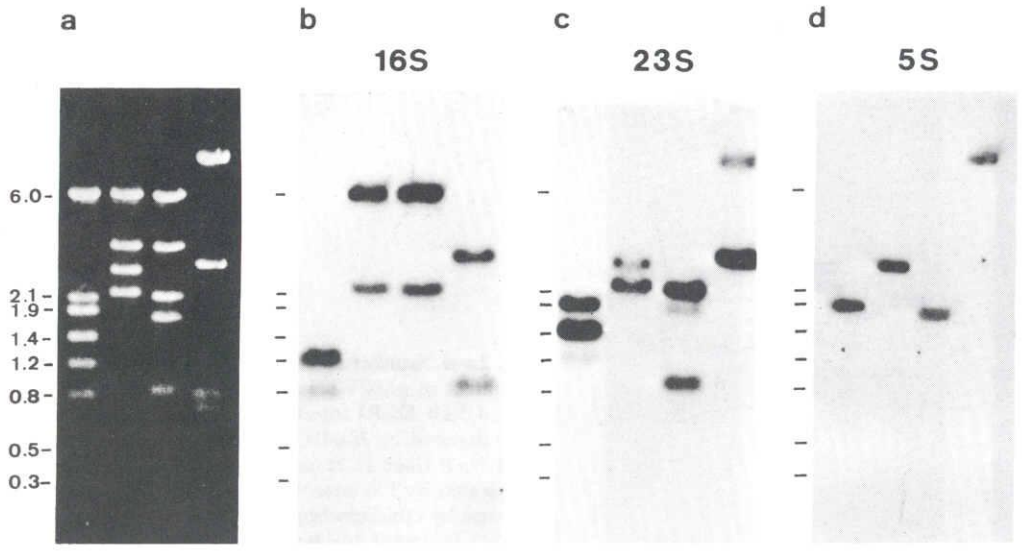



**Fig. 3a–d.** Southern hybridization of pMCB1002 DNA with $^{32}$P-labeled rRNAs. Total pMCB1002 DNA was digested by *Eco*RI/*Hin*dIII (lane 1), *Hpa*I (lane 2), *Hpa*I/*Sma*I (lane 3), and *Pst*I/*Eco*RI (lane 4), respectively. The products were separated by 1% agarose gel electrophoresis **a**, and hybridized with $^{32}$P-labeled 16S **b**, 23S **c** and 5S **d** rRNAs, respectively. The sizes (kb) of *Eco*RI/*Hin*dIII digestion products are shown in the left side of **a**

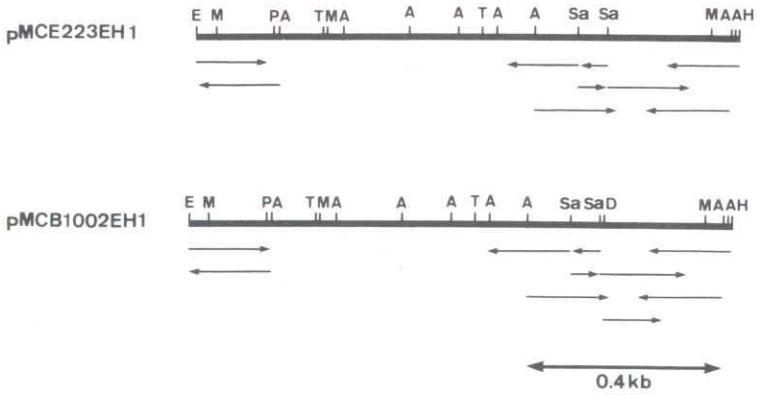



**Fig. 4.** Fine restriction maps of 1.2 kb DNA segments in pMCB1002EH1 and pMCE223EH1, and the sequencing strategy. "A" represents *Alu*I site. D: *Dra*I, H: *Hin*dIII, M: *Mbo*II, Sa: *Sau*3AI and T: *Tth*HB8I. The 0.4 kb *Alu*I-fragments, which were used for sequencing, are indicated by a thick horizontal arrow

*c) Nucleotide sequence of spacer region*

From the results in Figs. 1, 2 and 3, it is clear that the junction of the 16S and 23S rRNA structure genes falls in the 1.2 kb *Eco*RI/*Hin*dIII fragment of pMCE223 (*rrnA*) or pMCB1002 (*rrnB*). Thus, the 1.2 kb fragments from both plasmids were subcloned in pBR322 (pMCE223EH1 for *rrnA*, and pMCB1002EH1 for *rrnB*, respectively; Fig. 1c). The fragments hybridized strongly with 16S rRNA and weakly with 23S rRNA. Fine restriction maps of the two fragments were constructed, as shown in Fig. 4. The two maps were identical except that a *Dra*I cutting site existed only in pMCB1002EH1 (*rrnB*). The sequences of both terminal regions of the 1.2 kb DNA segments were first determined. About 90 nucleotides from the *Hin*dIII end were found to be highly homologous to the sequence of the 5′-terminus of 23S rRNA genes of *E. coli* (Brosius et al. 1981) and *Bacillus subtilis* (Loughney et al. 1982). Therefore, the 0.4 kb *Alu*I fragments including the putative 5′-terminal sequence of 23S rRNA gene and its upstream region (see Fig. 4) were isolated and sequenced according to the strategy shown in Fig. 4. Most of the sequences were established by sequencing both DNA strands more than twice. Figure 5 shows the sequences of the noncoding (RNA-like) DNA strand of the 0.4 kb *Alu*I fragments from both *rrnA* nad *rrnB*.The 3′-end of the 16S rRNA gene was deduced

```
rrnA  AGCTTAACCA TTTGGAGAGC GCTTCCCAAG GTAGGACTAG CGATTGGGGT GAAGTCGTAA  60
rrnB  ---------- ---------- ---------- ---------- ---------- ----------  60

      CAAGGTATCC GTACGGGAAC GTGCGGATTG ATCACCTCCT TTCTATGGAG ATATTTATAT 120
      ---------- ---------- --------G- ---------- ---------- ---------- 120
                                        16S<------┘
      TACTGACTAT TTAATTCTAT TTAGTTTTCA GAGATCGTAA AACATCTCTG AAAATACAGA 180
      ---------- ---------- ---------- --------C- C------T-A --T--**--- 178

      TTGTTCTTTG AAAACTGAAT ATTAGATGAA ATGCAATTTT CTGATTATAA CAATATTTAT 240
      ---------- ---------- ---------- ---------- ---------- ---------- 238

      AATTAGATAA TTATTACGAT ATTAAATTCG TAATGACATC AAAAACAATT AACTAAAATT 300
      ---------- ---------- ---------- -------T-- ---------- ---------- 298

      AATTGAGTTA CAAATTGCTA GAAAGATTTT CTAAAAAATA GTAAGAGCAT ATGGTGAATG 360
      ---------- ---------- ---------- ---------- ---------- ---------- 358
                                        └----------> 23S
      CCTTGGAAAA TGGAGCCGAA GAAGGACGTG ACTACCTGCG ATAAGTCTGG GGGAGCT    417
      ---------- ---------- ---------- ---------- ---------- -------    415
```

**Fig. 5.** Nucleotide sequences of the spacers between 16S and 23S rRNA genes. The total sequence of the 0.4 kb *Alu*I fragment from pMCE223EH1 (*rrnA*) is shown, compared with that from pMCB1002EH1 (*rrnB*). Bars indicate identical nucleotides and asterisks indicate deleted nucleotides

from the sequence of the 3′-terminal region of *M. capricolum* 16S rRNA (Woese et al. 1980). The 5′-end of the 23S rRNA gene was not determined experimentally but assigned on the basis of sequence homologies with *E. coli*, *B. subtilis* and *Anacystis nidulans* 23S rDNA (Brosius et al. 1981;

Loughney et al. 1982; Tomioka et al. 1981). Thus, the spacer between the 3′-end of 16S rRNA gene and the 5′-end of 23S rRNA gene in *rrnA* was estimated to be 228 nucleotides long, and that in *rrnB* was 226 nucleotides long. The two spacers revealed 97% sequence identity; only six nucleotide substitutions and two nucleotide insertions/deletions were found between them. The spacers were extremely poor in GC (20%), and did not include tRNA-like sequences. The sequences of the 3′-terminal region of the 16S rRNA and the 5′-terminal region of the 23S rRNA gene of *M. capricolum* so far determined revealed 69% (131/189) and 74% (139/189) identities with the corresponding sequences of *E. coli* and *B. subtilis*, respectively.

## Discussion

In a previous paper (Sawada et al. 1981), we demonstrated that there are at least two sets of closely linked genes for 16S, 23S, and 5S rRNA on the *M. capricolum* chromosome. The present study confirms the above observation. Two independent DNA segments carrying the genes for the three rRNA species have been cloned in the *E. coli* plasmids. The sizes of the cloned DNA fragments are coincident with those predicted by the genomic Southern hybridization (Sawada et al. 1981). Each of the two rRNA gene clusters, *rrnA* and *rrnB*, carries a set of genes for 16S, 23S, and 5S rRNAs. Southern hybridization analyses of the cloned DNAs and sequencings of the DNA segments have revealed that the order and orientation of the rRNA genes in each cluster are (5′)16S-23S-5S(3′). The structures of the two rRNA gene clusters of *M. capricolum* resemble each other as judged by the restriction maps. Although information on transcription of rRNA genes is lacking, it is plausible that each of the clusters forms a transcription unit. The order of the rRNA genes of *M. capricolum* is the same as that of other prokaryotes such as *E. coli* (Vola et al. 1977), *B. subtilis* (Stewart et al. 1982), *A. nidulans* (Tomioka et al. 1981) and *Rhodopseudomonas capsulata* (Yu et al. 1982).

Characteristic features of *M. capricolum* rRNA gene clusters can be found in the spacer regions between the 16S and 23S rRNA genes. First, the spacers do not include tRNA genes. All the seven rRNA operons of *E. coli* (Lund et al. 1976; Ikemura and Nomura 1977; Morgan et al. 1977), all the two rRNA gene clusters of *A. nidulans* (Tomioka et al. 1981) and two out of ten rRNA gene clusters of *B. subtilis* (Loughney et al. 1982) contain one or two spacer tRNA genes. Thus, the overall organization of the *M. capricolum* rRNA gene clusters resembles those of the eight *B. subtilis* clusters that do not contain spacer tRNA genes. Second, the spacer sequence of *M. capricolum* is extremely poor in GC (20%) in contrast to a relatively high GC-content (48%) of the coding sequence of rRNA genes (Ryan and Morowitz 1969). The GC-content of the spacer is even lower than that of the total genomic *M. capricolum* DNA (25%; Neimerk 1970). Third, the sequences of the two spacers of *M. capricolum* are highly conserved, revealing 97% sequence identity. Furthermore, most of the differences between the two sequence are found in a limited part of the spacer. This suggests that the bulk of the spacer sequence may play some functional roles. In *E. coli* and *B. subtilis*, the spacer sequences are relatively conserved among the different rRNA operons in the same species, while little sequence homology of the spacers has been found between the two species with exception of the spacer tRNA gene regions (Brosius et al. 1981; Loughney et al. 1982; Ogasawara et al. 1983). It has been suggested that the spacer sequence includes the processing signals of rRNA gene transcripts in the maturation steps of rRNA synthesis (Bram et al. 1980; Ogasawara et al. 1983). Interestingly, a remarkable sequence homology is found in certain regions of spacers and in the 5′-flanking sequences of the 16S rRNA genes between *M. capricolum* and *B. subtilis* (Iwami et al. 1984). These facts suggests that *M. capricolum* possesses a processing mechanism similar to that of *B. subtilis*. This will be discussed in detail in another paper (Iwami et al. 1984).

*Acknowledgements.* This work was supported by Grants from the Ministry of Education, Science and Culture, Japan (Nos. 56480377, 57121003, 57121009 and 58220012) and the Naito Science Foundation Research Grant.

## References

Bram RJ, Young RJ, Steitz JA (1980) The ribonuclease III site flanking 23S sequences in the 30S ribosomal precursor RNA of *E. coli*. Cell 19:393–401

Brosius J, Dull TJ, Sleeter DD, Noller HF (1981) Gene organization and primary structure of a ribosomal RNA operon from *Escherichia coli*. J Mol Biol 148:107–127

Grunstein M, Hogness DS (1975) Colony hybridization; a method for the isolation of cloned DNAs that contain a specific gene. Proc Natl Acad Sci USA 72:3961–3965

Hori H, Sawada M, Osawa S, Murano K, Ishikura H (1981) The nucleotide sequence of 5S rRNA from *Mycoplasma capricolum*. Nucl Acids Res 9:5407–5410

Ikemura T, Nomura M (1977) Expression of spacer tRNA genes in ribosomal RNA transcription units carried by hybrid Col El plasmids in E. coli. Cell 11:779–793

Iwami M, Muto A, Yamao F, Osawa S (1984) Nucleotide sequence of *rrnB* 16S ribosomal RNA gene from *Mycoplasma capricolum*. Mol Gen Genet 196:317–322

Kawauchi Y, Muto A, Osawa S (1982) The protein composition of *Mycoplasma capricolum*. Mol Gen Genet 188:7–11

Loughney K, Lund E, Dahlberg JE (1982) tRNA genes are found between the 16S and 23S rRNA genes in *Bacillus subtilis*. Nucl Acids Res 10:1607–1624

Lund E, Dahlberg JE, Lindahl L, Jaskunas SR, Dennis PP, Nomura M (1976) Transfer RNA genes between 16S and 23S rRNA genes in rRNA transcription units of *E. coli*. Cell 7:165–177

Maniloff J, Morowitz HJ (1972) Cell biology of the mycoplasmas. Bacteriol Rev 36:263–290

Maxam AM, Gilbert W (1977) A new method for sequencing DNA. Proc Natl Acad Sci USA 74:560–564

Morgan EA, Ikemura T, Nomura M (1977) Identification of spacer tRNA genes in individual ribosomal RNA transcription units of *Escherichia coli*. Proc Natl Acad Sci USA 74:2710–2714

Neimerk HC (1970) Division of mycoplasmas into subgroups. J Gen Microbiol 63:249–263

Ogasawara N, Moriya S, Yoshikawa H (1983) Structure and organization of rRNA operons in the region of the replication origin of the *Bacillus subtilis* chromosome. Nucl Acids Res 11:6301 6318

Oka A, Sugisaki H, Takanami M (1981) Nucleotide sequence of the kanamycin resistance transposon Tn903. J Mol Biol 147:217–226

Razin S (1978) The Mycoplasmas. Microbiol Rev 42:414–470

Ryan JL, Morowitz HJ (1969) Partial pruification of native rRNA and tRNA cistrons from *Mycoplasma* sp. (KID). Proc Natl Acad Sci USA 63:1282–1289

Sawada M, Osawa S, Kobayashi H, Hori H, Muto A (1981) The number of ribosomal RNA genes in *Mycoplasma capricolum*. Mol Gen Genet 182:502–504

Southern EM (1975) Detection of specific sequences among DNA fragments separated by gel electrophoresis. J Mol Biol 98:503–517

Stanbridge EJ, Reff EM (1979) The molceular biology of mycoplasmas. In: Barile MF, Razin S (eds) The Mycoplasmas 1, Academic Press, New York, p 157–185

Tomioka N, Shinozaki K, Sugiura M (1981) Molecular cloning and characterization of ribosomal RNA genes from a blue green alga, *Anacystis nidulans*. Mol Gen Genet 184:359–363

Vogelstein B, Gillespie D (1979) Preparative and analytical purification of DNA from agarose. Proc Natl Acad Sci USA 76:615–619

Vola C, Jarry B, Rosset R (1977) Linkage of 5S RNA and 16S+23S RNA genes on the *E. coli.* chromosome. Mol Gen Genet 153:337–341

Walseth TF, Johnson RA (1979) The enzymatic preparation of [gamma-$^{32}$P] nucleoside triphosphate, cyclic [$^{32}$P] AMP, and cyclic [$^{32}$P] GMP. Biochim Biophys Acta 562:11–31

Woese CR, Maniloff J, Zablen LB (1980) Phylogenetic analysis of the mycoplasmas. Proc Natl Acad Sci USA 77:494–498

Yu PL, John B, Falk H, Drews G (1982) Molecular cloning of the ribosomal RNA genes of the photosynthetic bacterium *Rhodopseudomonas capsulata*. Mol Gen Genet 188:392–398

Communicated by A. Böck

Received March 7 / April 10, 1984

Mol Gen Genet (1984) 196:317–322

MGG


# Nucleotide sequence of the *rrnB* 16S ribosomal RNA gene from *Mycoplasma capricolum*

**Masafumi Iwami*, Akira Muto, Fumiaki Yamao, and Syozo Osawa**

Laboratory of Molecular Genetics, Department of Biology, Faculty of Science, Nagoya University, Furo-Cho, Chikusa-Ku, Nagoya 464, Japan

**Summary.** The nucleotide sequences of the *rrnB* 16S ribosomal RNA gene and its 5′-and 3′-flanking regions from *Mycoplasma capricolum* have been determined. The coding sequence is 1521 base pairs long, being 21 base pairs shorter than that of the *Escherichia coli* 16S rRNA gene. The 16S rRNA sequence of *M. capricolum* reveals 74% and 76% identity with that of *E. coli* and *Anacystis nidulans*, respectively. The secondary structure model constructed from the *M. capricolum* 16S rRNA gene sequence resembles that proposed for *E. coli* 16S rRNA. A large stem structure can be constructed between the 5′- and 3′-flanking sequences of the 16S rRNA gene. The flanking regions are extremely rich in AT.

## Introduction

*Myoplasma capricolum* contains only two sets of rRNA genes in the genome (Sawada et al. 1981). In a previous paper, we have shown that the rRNA genes from *M. capricolum* are clustered and arranged in the order 16S-23S-5S (Sawada et al. 1984). Thus the organization of the rRNA gene clusters of *M. capricolum* is essentially the same as that of other prokaryotes. However, in contrast to other prokaryotes, the *M. capricolum* rRNA gene clusters have no tRNA genes in the spacer between 16S and 23S rRNA genes (Sawada et al. 1984). It has been also known that 5S rRNA of *M. capricolum* is only 107 nucleotides long, the shortest of all the 5S rRNAs so far known (Hori et al. 1981). The 16S rRNA of *M. capricolum* has also been reported to be shorter than other bacterial 16S rRNAs (Stanbridge and Reff 1979). Furthermore, the GC-contents of the *Mycoplasma* rRNAs are the lowest among the prokaryotes (Ryan and Morowitz 1969). These data suggest that the *M. capricolum* rRNA has unique characteristics as compared with other bacterial rRNA.

In this study, we have determined the nucleotide sequences of a 16S rRNA gene from *M. capricolum* and its flanking regions and compared them with those of prokaryotic 16S rRNA genes from *Escherichia coli* (Brosius et al. 1978; Carbon et al. 1978, 1979) and a blue-green alga, *Anacystis nidulans* (Tomioka and Sugiura 1983).

* *Present address:* Meiji Institute of Health Science, 540 Naruda, Odawara 250, Japan

*Offprint requests to:* A. Muto

## Materials and methods

The recombinant plasmid pMCH158 contains one of the 16S rRNA genes and its 5′- and 3′-flanking regions of *M. capricolum* (Sawada et al. 1984). The plasmids were propagated in *E. coli* HB101 cells and isolated by the procedure of Oka et al. (1981). The DNA was digested by *Hin*dIII and separated by 0.7% agarose gel electrophoresis. The 2.0 kilo bases (kb) insert DNA fragment was eluted from the excised gel pieces using glass beads (Vogelstein and Gilespie 1979). The 5′-terminal end-labeling was carried out by using ($\gamma$-$^{32}$P)ATP and T4 polynucleotide kinase after alkaline phosphatase treatment as described by Maxam and Gilbert (1980). A fine structure map of the insert DNA was constructed by the method of Smith and Birnstiel (1976). Polyacrylamide gels (5%) were used for separating labeled fragments and for the strand separations. When the strand separation was unsuccessful, the labeled fragment was digested by the second restriction endonuclease treatment and separated by gel electrophoresis. The DNA fragments were eluted from polyacrylamide gels by diffusion, and base-specific chemical cleavages (G, G+A, T+C, C) were performed according to Maxam and Gilbert (1980).

Other materials and methods were as described (Sawada et al. 1984).

## Results and discussion

The plasmid pMCH158 contained a 2.0 kb *Hin*dIII *M. capricolum* DNA fragment which hybridized with 16S and weakly with 23S rRNA. Figure 1(a) shows a fine structure map of the 2.0 kb insert DNA. Most parts of the map were identical with a part of the map of pMCB1002 which contained the *rrnB* rRNA gene cluster of *M. capricolum* (see Fig. 4 of Sawada et al. 1984). Since the locations of rRNA genes in pMCB1002 are known, the 2.0 kb fragment in pMCH158 is expected to contain a full length of the 16S rRNA gene and the 5′-terminal region of the 23S rRNA gene of the *rrnB* cluster. Thus, the complete sequence of the 2.0 kb fragment was determined. The sequencing strategy is outlined in Fig. 1(b). Most of the sequences were confirmed by sequencing both DNA strands.

Figure 2 shows the total nucleotide sequence of the noncoding (RNA-like) strand of the 2.0 kb insert of pMCH158. The 3′-end was assigned by the sequence of the 3′-terminal region of *M. capricolum* 16S rRNA reported by Woese et al. (1980a). The 5′-end was assigned on the bases of sequence

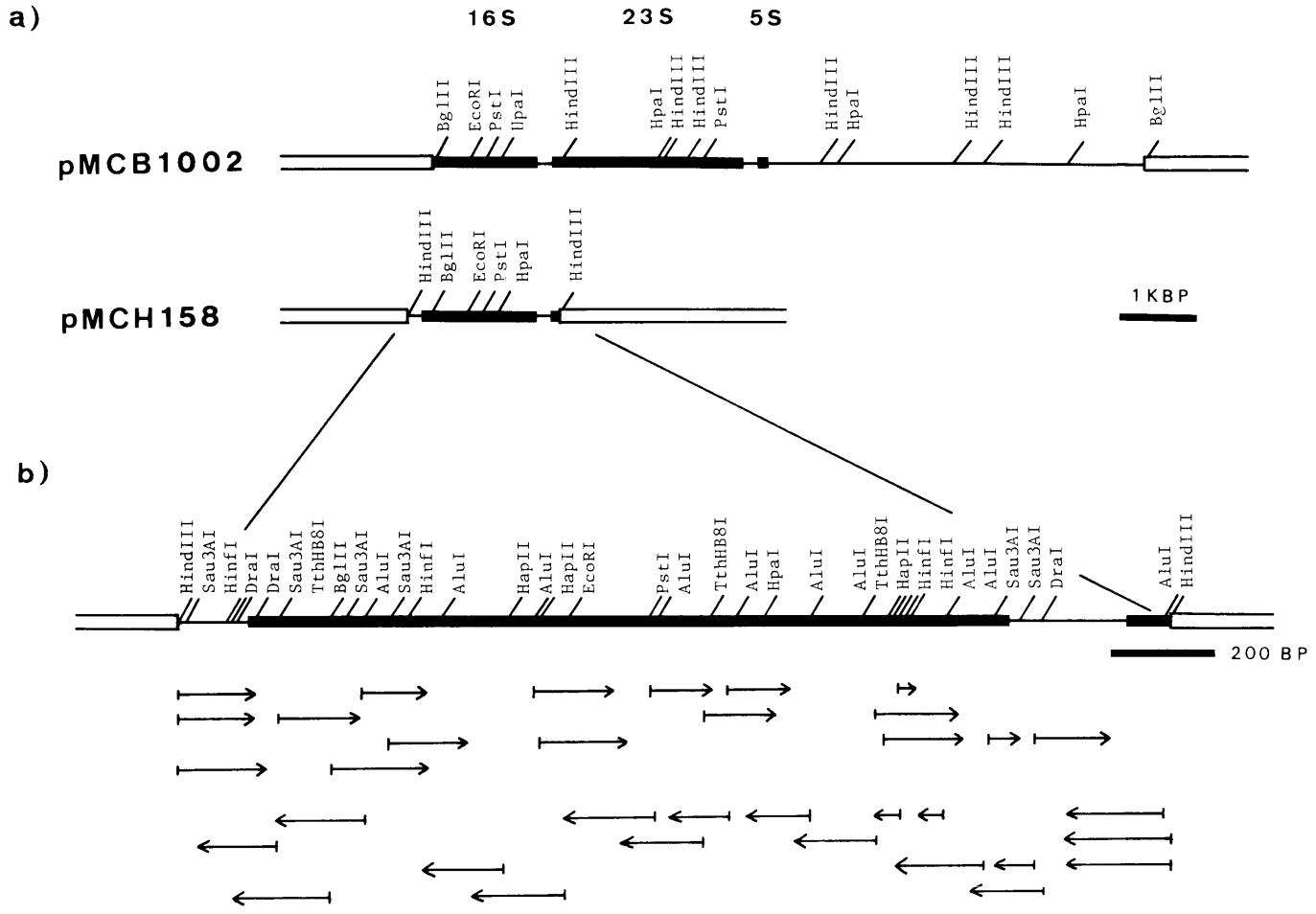



**Fig. 1.** Physical maps of the cloned rRNA gene cluster in plasmids pMCB1002 and pMCH158 (**a**) and fine structure map and sequencing strategy for the insert of plasmid pMCH158 (**b**). Thick lines represent the rRNA coding regions. The vector pBR322 DNA is shown by open boxes. The lower part shows the sequencing strategy. *Arrows* indicate the direction and length of the DNA regions sequenced. KBP: kilo-base pairs. BP: base pairs

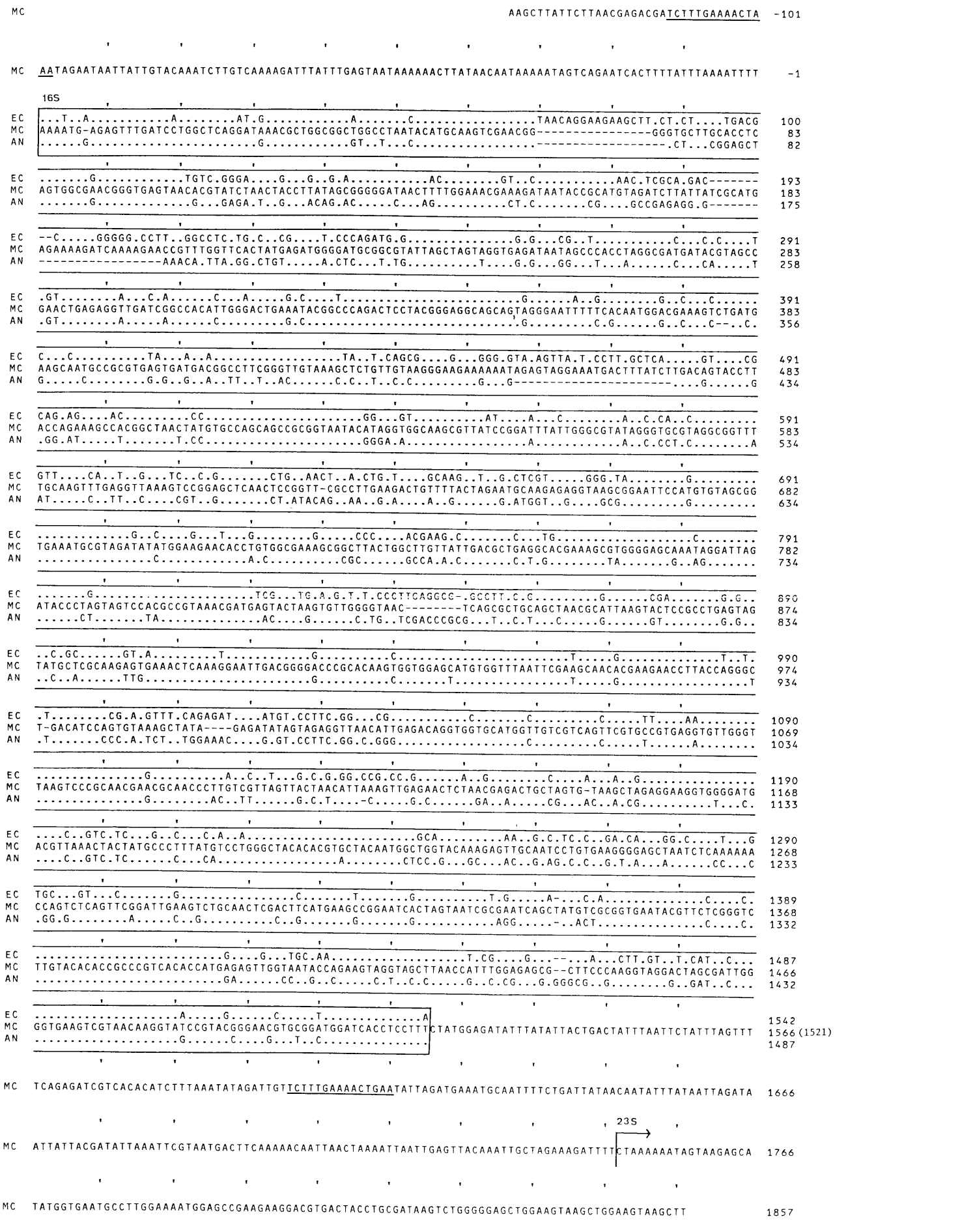
homology with that of *E. coli* (Brosius et al. 1978) and *A. nidulans* (Tomioka and Sugiura 1983). Thus, the coding region of the *M. capricolum* 16S rRNA gene was 1521 base pairs (bp) long. As shown in the figure, the 2.0 kb insert also contained the 5′-flanking sequence (134 bp) of the 16S rRNA gene, the spacer between the 16S and 23S rRNA genes (226 bp) and the 5′-terminal sequence of the 23S rRNA gene (110 bp) (see also Sawada et al. 1984). In Fig. 2, the sequence of the 16S rRNA gene of *M. capricolum* was aligned together with that of *E. coli* (Brosius et al. 1978) and *A. nidulans* (Tomioka and Sugiura 1983). The *M. capricolum* sequence revealed a remarkable similarity to those of *E. coli* (74% identity) and *A. nidulans* (76% identity).

The identity values were higher than those for the 5S rRNA between *M. capricolum* and *E. coli* or *A. nidulans* (56% or 53%; Hori et al. 1981; Dyer and Bowman 1979). The GC content of the *M. capricolum* 16S rRNA gene was 48%, which was lower than that of *E. coli* (55%) and *A. nidulans* (56%).

The 16S rRNA gene of *M. capricolum* was 21 bp shorter than that of *E. coli* and 34 bp longer than that of *A. nidulans*. When compared with *E. coli*, the *M. capricolum* gene contained three relatively large deletions (positions 68–69, 833–834 and 996–997) and an insertion (position 177–185). The deletion at position 68–69 also existed in the corresponding region of the *A. nidulans* 16S rRNA gene. The sequence homology around these deletion and insertion regions was low among the three species. On the other hand, highly conserved regions were: (i) the 5′-terminal region, (ii) position 498–546 which corresponds to the proposed ribosomal protein S4 binding site in *E. coli* (Nomura et al. 1980), and (iii) the 3′-terminal region containing the sequence 5′-ACCTCC-3′, a proposed prokaryotic mRNA binding sequence (Shine and Dalgarno 1974).

Secondary structure models for *E. coli* 16S rRNA have been proposed by three groups (Woese et al. 1980b; Stiegler et al. 1981; Zwieb et al. 1981). Two of these models have been revised recently (Maly and Brimacombe 1983; Woese et al. 1983), thus reaching essential agreement between them. The secondary structure of *M. capricolum* 16S rRNA

**Fig. 2.** Total nucleotide sequence of the insert of pMCH158. The noncoding (RNA-like) strand of *M. capricolum* (MC) 16S rRNA is shown together with that of *E. coli* (EC) and *A. nidulans* (AN) 16S rRNA genes for comparison. The 16S rRNA coding regions are boxed. Dots indicate nucleotides identical with those of *M. capricolum*. Bars indicate deletions. Repeated sequences are underlined. Numbering in the right-hand margin is for each 16S rRNA species


MC AAGCTTATTCTTAACGAGACGATCTTTGAAAACTA -101
MC AATAGAATAATTATTGTACAAATCTTGTCAAAAGATTTATTTGAGTAATAAAAAACTTATAACAATAAAAATAGTCAGAATCACTTTTATTTAAAATTTT -1
16S
EC ...T..A.....A.....AT.G.....A.....C.....TAACAGGAAGAAGCTT.CT.CT....TGACG 100
MC AAAATG-AGAGTTTGATCCTGGCTCAGGATAAACGCTGGCGGCTGGCCTAATACATGCAAGTCGAACGG----------GGGTGCTTGCACCTC 83
AN ......G.....G.....GT..T...C.....----------.CT...CGGAGCT 82
EC .....G.....TGTC.GGGA....G...G..G.A.....AC.....GT..C.....AAC.TCGCA.GAC------- 193
MC AGTGGCGAACGGGTGAGTAACACGTATCTAACTACCTTATAGCGGGGGATAACTTTTGGAAACGAAAGATAATACCGCATGTAGATCTTATTATCGCATG 183
AN .....G.....G...GAGA.T..G...ACAG.AC.....C...AG.....CT.C.....CG....GCCGAGAGG.G------- 175
EC --C.....GGGGG.CCTT..GGCCTC.TG.C..CG....T.CCCAGATG.G.....G.G...CG..T.....C...C.C....T 291
MC AGAAAAGATCAAAAGAACCGTTTGGTTCACTATGAGATGGGGATGCGGCGTATTAGCTAGTAGGTGAGATAATAGCCCACCTAGGCGATGATACGTAGCC 283
AN ----------AAACA.TTA.GG.CTGT.....A.CTC...T.TG.....T....G.G...GG...T...A.....C...CA.....T 258
EC .GT.....A...C.A.....C...A.....G.C....T.....G.....A..G.....G..C...C..... 391
MC GAACTGAGAGGTTGATCGGCCACATTGGGACTGAAATACGGCCCAGACTCCTACGGGAGGCAGCAGTAGGGAATTTTTCACAATGGACGAAAGTCTGATG 383
AN .GT.....A.....A.....C.....G.C.....G.....C.G.....G..C...C--..C. 356
EC C...C.....TA...A..A.....TA..T.CAGCG....G...GGG.GTA.AGTTA.T.CCTT.GCTCA.....GT....CG 491
MC AAGCAATGCCGCGTGAGTGATGACGGCCTTCGGGTTGTAAAGCTCTGTTGTAAGGGAAGAAAAAATAGAGTAGGAAATGACTTTATCTTGACAGTACCTT 483
AN G.....C.....G.G..G..A..TT..T..AC.....C.C..T..C.C.....G...G----------....G.....G 434
EC CAG.AG....AC.....CC.....GG...GT.....AT....A...C.....A..C.CA..C..... 591
MC ACCAGAAAGCCACGGCTAACTATGTGCCAGCAGCCGCGGTAATACATAGGTGGCAAGCGTTATCCGGATTTATTGGGCGTATAGGGTGCGTAGGCGGTTT 583
AN .GG.AT.....T.....T.CC.....GGGA.A.....A.....A..C.CCT.C.....A 534
EC GTT....CA..T..G...TC..C.G.....CTG..AACT..A.CTG.T....GCAAG..T..G.CTCGT.....GGG.TA.....G..... 691
MC TGCAAGTTTGAGGTTAAAGTCCGGAGCTCAACTCCGGTT-CGCCTTGAAGACTGTTTTACTAGAATGCAAGAGAGGTAAGCGGAATTCCATGTGTAGCGG 682
AN AT.....C..TT..C....CGT..G.....CT.ATACAG..AA..G.A....A..G.....G.ATGGT..G....GCG.....G..... 634
EC .....G..C....G...T...G.....G.....CCC....ACGAAG.C.....C...TG.....C..... 791
MC TGAAATGCGTAGATATATGGAAGAACACCTGTGGCGAAAGCGGCTTACTGGCTTGTTATTGACGCTGAGGCACGAAAGCGTGGGGAGCAAATAGGATTAG 782
AN .....C.....A.C.....CGC.....GCCA.A.C.....C.T.G.....TA.....G..AG..... 734
EC .....G.....TCG...TG.A.G.T.T.CCCTTCAGGCG-.GCCTT.C.G.....G.....CGA.....G.G.. 890
MC ATACCCTAGTAGTCCACGCCGTAAACGATGAGTACTAAGTGTTGGGGTAAC--------TCAGCGCTGCAGCTAACGCATTAAGTACTCCGCCTGAGTAG 874
AN .....CT.....TA.....AC....G.....C.TG..TCGACCCGCG...T..C.T...C.....G.....GT.....G.G.. 834
EC ..C.GC.....GT.A.....T.....G.....C.....T.....G.....T..T. 990
MC TATGCTCGCAAGAGTGAAACTCAAAGGAATTGACGGGGACCCGCACAAGTGGTGGAGCATGTGGTTTAATTCGAAGCAACACGAAGAACCTTACCAGGGC 974
AN ..C..A.....TTG.....G.....C.....T.....T.....G.....T 934
EC .T.....CG.A.GTTT.CAGAGAT....ATGT.CCTTC.GG...CG.....C.....C.....C.....TT....AA..... 1090
MC T-GACATCCAGTGTAAAGCTATA----GAGATATAGTAGAGGTTAACATTGAGACAGGTGGTGCATGGTTGTCGTCAGTTCGTGCCGTGAGGTGTTGGGT 1069
AN .T.....CCC.A.TCT..TGGAAAC....G.GT.CCTTC.GG.C.GGG.....C.....C.....T.....A..... 1034
EC .....G.....A..C..T...G.C.G.GG.CCG.CC.G.....A..G.....C....A...A..G..... 1190
MC TAAGTCCCGCAACGAACGCAACCCTTGTCGTTAGTTACTAACATTAAAGTTGAGAACTCTAACGAGACTGCTAGTG-TAAGCTAGAGGAAGGTGGGGATG 1168
AN .....G.....AC..TT.....G.C.T....-C.....G.C.....GA..A.....CG...AC..A.CG.....T...C. 1133
EC ....C..GTC.TC...G..C...C.A..A.....GCA.....AA..G.C.TC.C..GA.CA...GG.C....T...G 1290
MC ACGTTAAACTACTATGCCCTTTATGTCCTGGGCTACACACGTGCTACAATGGCTGGTACAAAGAGTTGCAATCCTGTGAAGGGGAGCTAATCTCAAAAAA 1268
AN ....C..GTC.TC.....C...CA.....A.....CTCC.G...GC...AC..G.AG.C.C..G.T.A...A.....CC...C 1233
EC TGC...GT...C.....G.....C.....T.....G.....T.G.....A-...C.A.....C....C. 1389
MC CCAGTCTCAGTTCGGATTGAAGTCTGCAACTCGACTTCATGAAGCCGGAATCACTAGTAATCGCGAATCAGCTATGTCGCGGTGAATACGTTCTCGGGTC 1368
AN .GG.G.....A.....C..G.....C..G.....G.....G.....AGG.....-..ACT.....C....C. 1332
EC .....G....G...TGC.AA.....T.CG....G...--...A...CTT.GT..T.CAT..C.... 1487
MC TTGTACACACCGCCCGTCACACCATGAGAGTTGGTAATACCAGAAGTAGGTAGCTTAACCATTTGGAGAGCG--CTTCCCAAGGTAGGACTAGCGATTGG 1466
AN .....GA.....CC..G..C.....C.T..C.C.....G..C.CG...G.GGGCG..G.....G..GAT..C.... 1432
EC .....A.....G.....C.....T.....A 1542
MC GGTGAAGTCGTAACAAGGTATCCGTACGGGAACGTGCGGATGGATCACCTCCTTTCTATGGAGATATTTATATTACTGACTATTTAATTCTATTTAGTTT 1566 (1521)
AN .....G.....C....G...T..C..... 1487
MC TCAGAGATCGTCACACATCTTTAAATATAGATTGTTCTTTGAAAACTGAATATTAGATGAAATGCAATTTTCTGATTATAACAATATTTATAATTAGATA 1666
23S
MC ATTATTACGATATTAAATTCGTAATGACTTCAAAAACAATTAACTAAAATTAATTGAGTTACAAATTGCTAGAAAGATTTTCTAAAAAATAGTAAGAGCA 1766
MC TATGGTGAATGCCTTGGAAAATGGAGCCGAAGAAGGACGTGACTACCTGCGATAAGTCTGGGGGAGCTGGAAGTAAGCTGGAAGTAAGCTT 1857

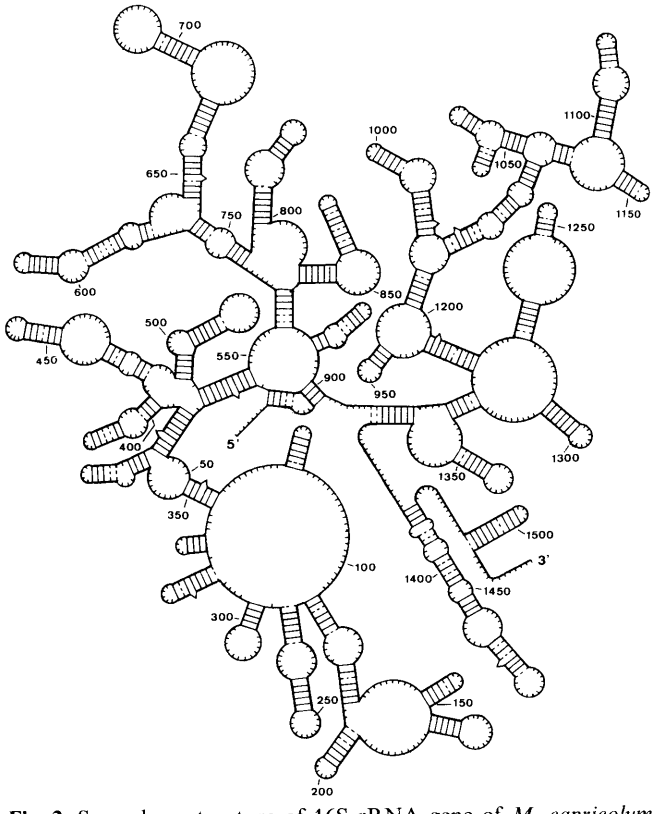



**Fig. 3.** Secondary structure of 16S rRNA gene of *M. capricolum*. Crossbars indicate normal base pairs and dotted bars represent G–U pairs

constructed according to them was nearly identical to that of *E. coli* as shown in Fig. 3. The sequence identity between *E. coli* and *M. capricolum* 16S rRNAs was 74%, and yet almost all of the stem and loop structures in *E. coli* 16S rRNA could be observed in *M. capricolum*, indicating a strong conservation of the secondary structures between these two bacterial 16S rRNAs. The sequence conservation is greater in nonpaired (loop) regions than in paired (stem) regions. This means that a base change occurring in the stem regions has been cancelled by a compensatory change of the opposite base, so that the secondary structure can be maintained. A striking example of secondary structure conservation is seen in the stem and loop structures of positions 67–86 and 809–866 of *M. capricolum* compared with the corresponding regions of *E. coli*, although the sequences of these regions were not well conserved between the two 16S rRNAs with deletions in *M. capricolum* (Fig. 4). On the contrary, there were several regions where the primary as well as the secondary structures did not resemble each other. A stem of 8 bp long between the positions 588–595 and 644–651 in *E. coli* has been suggested as a part of ribosomal proteins S8 and S15 binding sites (Ungewickell et al. 1975; Zimmermann et al. 1975; Zimmermann et al. unpublished results), while the corresponding region in *M. capricolum* (positions 580–587 and 636–643) had only three base pairs (positions 636–638 vs 585–587), forming a fairly large loop with the rest of the bases in this region (Fig. 5).

The 5′-flanking region as well as the spacer between 16S and 23S rRNA genes was extremely poor in GC (20%) (Fig. 2), having no sequence homology with those of *E. coli*. The homology of the flanking sequences between *M.*

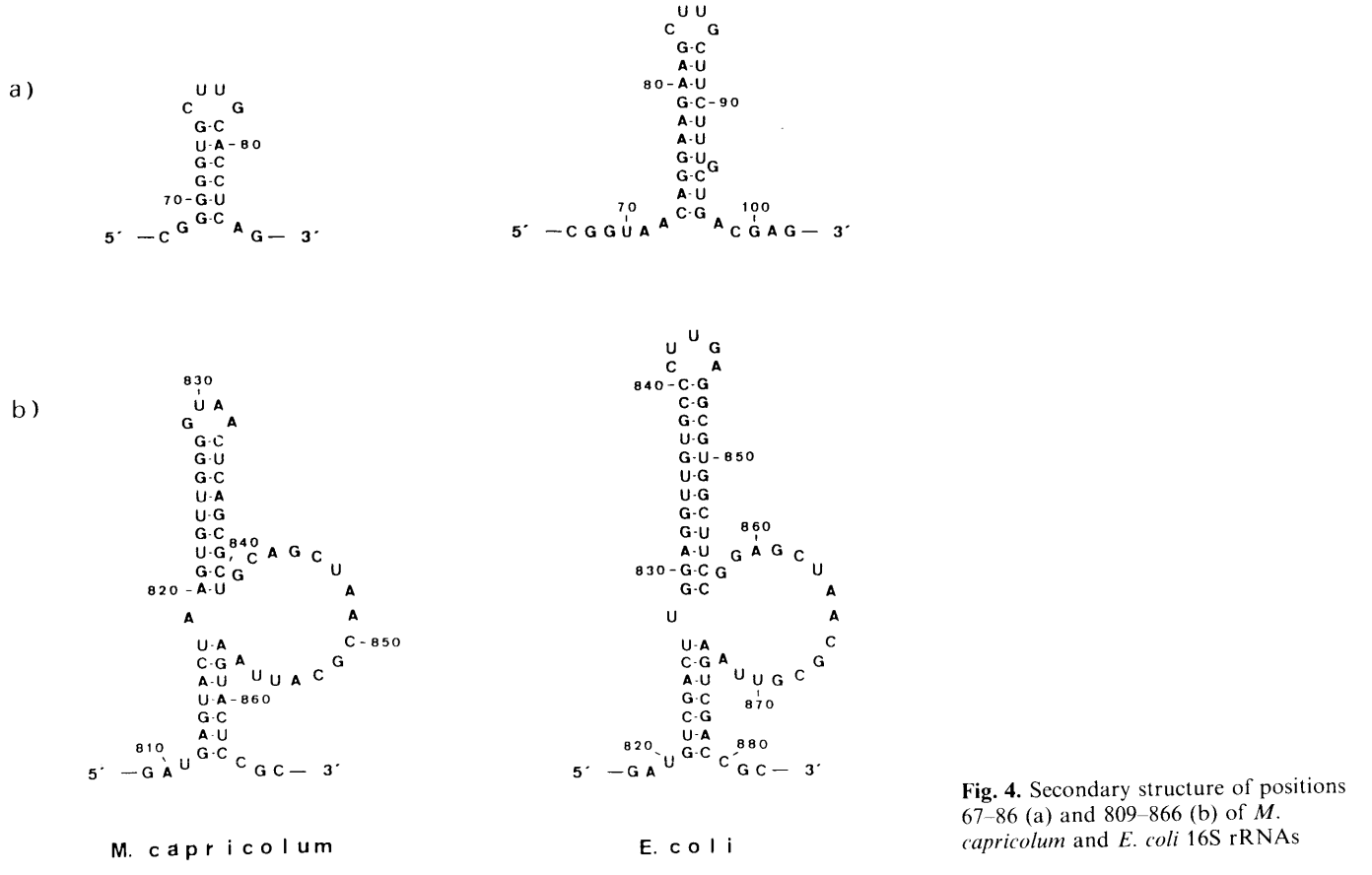



**Fig. 4.** Secondary structure of positions 67–86 (a) and 809–866 (b) of *M. capricolum* and *E. coli* 16S rRNAs

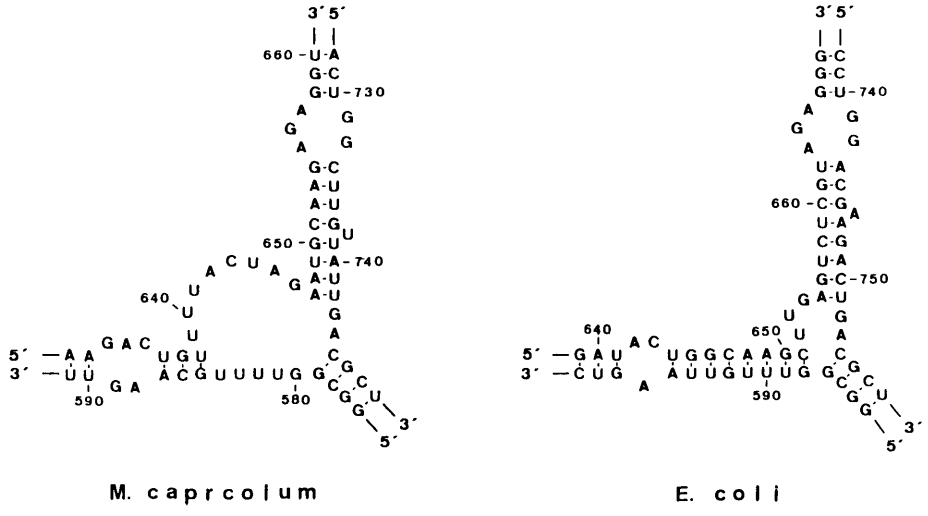



**Fig. 5.** Secondary structure of positions (576–591, 631–660, and 728–748) of *M. capricolum* and *E. coli* 16S rRNAs

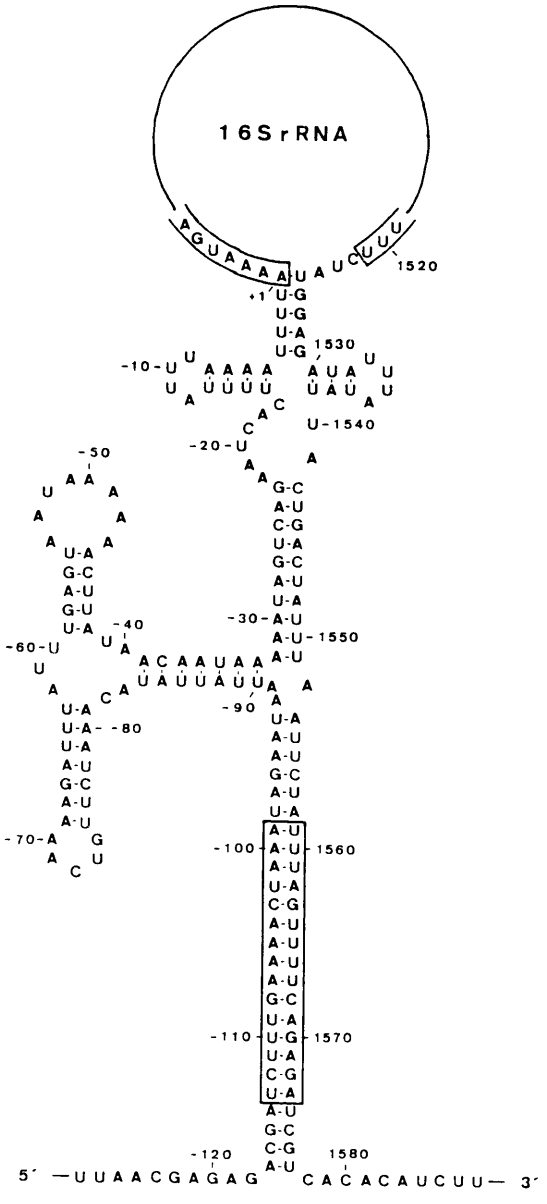



**Fig. 6.** Secondary structure of the regions surrounding the 5′- and 3′-ends of *M. capricolum* 16S rRNA. Homologous stretches between *M. capricolum* and *B. subtilis* are boxed

*capricolum* and *Bacillus subtilis* was also very low with, however, certain parts highly conserved (see later). No promoter-like sequence was found in the 5′-flanking region, because this insert was not long enough to include the promoter region. The spacer sequence between the 16S and 23S rRNA genes was identical with that in pMCB1002 (Sawada et al. 1984), confirming that the rRNA genes inserted in pMCH158 were a part of the *rrnB* cluster.

A possible secondary structure model of the flanking sequences of the 16S rRNA gene was constructed (Fig. 6). A large stem structure could be generated by pairing the 5′-flanking sequences of the 16S rRNA gene with the 5′-terminal region of the 16S and 23S spacer as in the case of *E. coli* (Young and Steitz 1978) and in *B. subtilis* (Ogasawara et al. 1983; Stewart and Bott 1983). These stems have been suggested as possible substrates for processing enzymes during rRNA maturation (Young and Steitz 1978; Ogasawara et al. 1983). Furthermore, a stretch having almost the same sequence as position −111––98 appeared again in the spacer at position 1602–1616 (see underlined sequences in Fig. 2), which can base pair with a part of the spacer sequence between the 23S and 5S rRNA genes (unpublished result). Thus, there existed two almost identical base-paired structures in the 16S rRNA stem and in the 23S rRNA stem. These repeated structures in the two stems have also been reported in *B. subtilis* by Ogasawara et al. (1983), but not in *E. coli*. Moreover, a striking homology between *M. capricolum* and *B. subtilis* was found in the sequences including the repeated stem regions. As shown in Fig. 7, a stretch in the 5′-flanking sequence (position −116––77) and two stretches in the 16S–23S spacer sequence (position 1557–1577 and 1601–1620) were highly homologous with those of the corresponding regions of *B. subtilis*, but not with those of *E. coli*. The repeated structures in the two stems were located within the homologous regions as compared with *B. subtilis* (the sequence boxed in Fig. 6). Ogasawara et al. (1983) have suggested that these regions may be the target sites for processing enzymes. This would mean that the processing signals of rRNA gene transcripts as well as the specificity of processing enzymes are conserved between *M. capricolum* and *B. subtilis*, but not between these two species and *E. coli*. We have previously demonstrated that the mycoplasmas are phylogenetically more closely related to the gram-positive bacteria such as *B. subtilis* rather than gram-negative bacteria such as *E.*

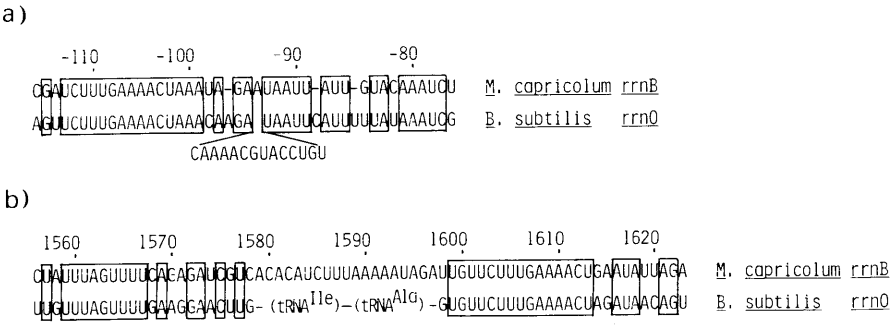



**Fig. 7.** Sequence comparisons (a) in the 5′-upstream regions of 16S rRNAs and (b) in the spacer regions between *M. capricolum* (*rrnB*) and *B. subtilis* (*rrnO*) (Ogasawara et al. 1983). Sequences are aligned to obtain maximum homology. Identical sequences are boxed.

*coli*, from the 5S rRNA sequence (Hori et al. 1981) and ribosomal proteins (Kawauchi et al. 1982). The similarity of the spacer structures of the rRNA gene clusters between *M. capricolum* and *B. subtilis* also supports this conclusion.

*Acknowledgements.* We thank Dr. H. Hori of this laboratory for his help on the computer analyses. This work was supported by Grants from the Ministry of Education, Science and Culture (Nos. 56480377, 57121003, 57121009 and 58220012) and the Naito Science Foundation Research Grant.

## References

Brosius J, Palmer ML, Kennedy PJ, Noller HF (1978) Complete nucleotide sequence of a 16S ribosomal RNA gene from *Escherichia coli*. Proc Natl Acad Sci USA 75:4801–4805

Brosius J, Dull TJ, Sleeter DD, Noller HF (1981) Gene organization and primary structure of a ribosomal RNA operon from *Escherichia coli*. J Mol Biol 148:107–127

Carbon P, Ehresmann C, Ehresmann B, Ebel JP (1978) The sequence of *Escherichia coli* ribosomal 16S RNA determined by new gel methods. FEBS Lett 94:152–156

Carbon P, Ehresmann C, Ehresmann B, Ebel JP (1979) The complete nucleotide sequence of the ribosomal 16-S RNA from *Escherichia coli*: Experimental details and cistron heterogeneities. Eur J Biochem 100:399–410

Dyer TA, Bowman CM (1979) Nucleotide sequences of chloroplast 5S ribosomal ribonucleic acied in flowering plants. Biochem J 183:595–604

Hori H, Sawada M, Osawa S, Murao K, Ishikura H (1981) The nucleotide sequence of 5S rRNA from *Mycoplasma capricolum*. Nucl Acids Res 9:5407–5410

Kawauchi Y, Muto A, Osawa S (1982) The protein composition of *Mycoplasma capricolum*. Mol Gen Genet 188:7–11

Maly P, Brimacombe R (1983) Refined secondary structure models for the 16S and 23S ribosomal RNA of *Escherichia coli*. Nucl Acids Res 11:7263–7286

Maxam AM, Gilbert W (1980) Sequencing end-labeled DNA with base-specific chemical cleavages. Methods Enzymol 65:499–460

Nomura N, Yates JL, Dean D, Post LE (1980) Feedback regulation of ribosomal protein gene expression in *Escherichia coli*: Structural homology of ribosomal RNA and ribosomal protein mRNA. Proc Natl Acad Sci USA 77:7084–7088

Oka A, Sugisaki H, Takanami M (1981) Nucleotide sequence of the kanamycin resistance transposon Tn*903*. J Mol Biol 147:217–226

Ogasawara N, Moriya S, Yoshikawa H (1983) Structure and organization of rRNA operons in the region of the replication origin of the *Bacillus subtilis* chromosome. Nucl Acid Res 11:6301–6318

Ryan JL, Morowitz HJ (1969) Partial purification of native rRNA and tRNA cistron from *Mycoplasma* sp. (KID). Proc Natl Acad Sci USA 63:1282–1289

Sawada M, Osawa S, Kobayashi H, Hori H, Muto A (1981) The number of ribosomal RNA genes in *Mycoplasma capricolum*. Mol Gen Genet 182:502–504

Sawada M, Muto A, Iwami M, Yamao F, Osawa S (1984) Organization of ribosomal RNA gene in *Mycoplasma capricolum*. Mol Gen Genet 196:311–316

Smith HO, Birnstiel ML (1976) A simple method for DNA restriction site mapping. Nucl Acids Res 3:2387–2398

Shine J, Dalgarno L (1974) The 3′-terminal sequence of *Escherichia coli* 16S ribosomal RNA: Complementarity to nonsense triplets and ribosomal binding sites. Proc Natl Acad Sci USA 71:1342–1346

Stanbridge EJ, Reff ME (1979) The molecular biology of mycoplasmas. In: Barile MF, Razin S (eds) The Mycoplasmas 1, Academic Press, New York, p 157–185

Stewart GC, Bott KF (1983) DNA sequence of the ribosomal RNA promoter for *B. subtilis* operon *rrnB*. Nucl Acids Res 11:6289–6300

Stiegler P, Carbon P, Zuker M, Ebel JP, Ehresmann C (1981) Structural organization of the 16S ribosomal RNA from *E. coli*: Topography and secondary structure. Nucl Acids Res 9:2153–2172

Tomioka N, Sugiura M (1983) The complete nucleotide sequence of a 16S ribosomal RNA gene from a blue-green alga, *Anacystis nidulans* Mol Gen Genet 191:46–50

Ungewickell E, Garrett R, Ehresmann C, Stiegler P, Fellner P (1975) An investigatin of the 16-S RNA binding sites of ribosomal proteins S4, S8, S15 and S20 from *Escherichia coli*. Eur J Biochem 51:165–180

Vogelstein B, Gillespie D (1979) Preparative and analytical purification of DNA from agarose. Proc Natl Acads Sci USA 76:615–619

Woese CR, Maniloff J, Zablen LB (1980a) Phylogenetic analysis of the mayoplasmas. Proc Natl Acad Sci USA 77:494–498

Woese CR, Magrum LT, Gupta R, Siegel RB, Stahl DA, Kop J, Crawford N, Brosius J, Gutell R, Hogan JJ, Noller H(1980b) Secondary structure model for bacterial 16S ribosomal RNA: Phylogenetic, enzymatic and chemical evidence. Nucl Acids Res 8:2275–2283

Woese CR, Gutell R, Gupta R, Noller HF (1983) Detailed analysis of the higher-order structure of 16S-like ribosomal ribonucleic acids. Microbiol Rev 47:621–669

Young RA, Steitz JA (1978) Complementary sequences 1700 nucleotides apart from a ribonuclease III cleavage site in *Escherichia coli* ribosomal precursor RNA. Proc Natl Acad Sci USA 75:3593–3597

Zimmermann RA, Mackie G, Muto A, Garrett R, Ungewickell E, Ehresmann C, Stiegler P, Ebel JP (1975) Location and characteristics of ribosomal protein binding sites in the 16S RNA of *Escherichia coli*. Nucl Acid Res 2:279–302

Zwieb C, Glotz C, Brimacombe R (1981) Secondary structure comparisons between small subunit ribosomal RNA molecules from six different species. Nucl Acid Res 9:3621–3640

Communicated by A. Böck

Received March 7, 1984

*Proc. Natl. Acad. Sci. USA*
Vol. 82, pp. 2306–2309, April 1985
Biochemistry

# UGA is read as tryptophan in *Mycoplasma capricolum*

(ribosomal proteins/opal tRNA/*Mycoplasma* genetic code/*Mycoplasma* tRNA)

FUMIAKI YAMAO, AKIRA MUTO, YASUSHI KAWAUCHI*, MASAFUMI IWAMI†, SHOJI IWAGAMI, YOSHITAKA AZUMI, AND SYOZO OSAWA

Laboratory of Molecular Genetics, Department of Biology, Faculty of Science, Nagoya University, Chikusa-ku, Nagoya 464, Japan

*Communicated by Motoo Kimura, December 11, 1984*

**ABSTRACT** **UGA is a nonsense or termination (opal) codon throughout prokaryotes and eukaryotes. However, mitochondria use not only UGG but also UGA as a tryptophan codon. Here, we show that UGA also codes for tryptophan in *Mycoplasma capricolum*, a wall-less bacterium having a genome only 20–25% the size of the *Escherichia coli* genome. This conclusion is based on the following evidence. First, the nucleotide sequence of the S3 and L16 ribosomal protein genes from *M. capricolum* includes UGA codons in the reading frames; they appear at positions corresponding to tryptophan in *E. coli* S3 and L16. Second, a $tRNA^{Trp}$ gene and its product tRNA found in *M. capricolum* have the anticodon sequence 5′ U-C-A 3′, which can form a complementary base-pairing interaction with UGA.**

We recently have sequenced a part of the *Mycoplasma capricolum* ribosomal-protein gene cluster that codes for polypeptides highly homologous to the *Escherichia coli* ribosomal proteins S3 and L16. The sequence contains four UGA codons in the reading frames; three appear at the sites corresponding to tryptophan, and one, at a site corresponding to arginine in the *E. coli* proteins. No "universal" UGG codon for tryptophan has so far been found. We have also isolated a clone containing a pair of *M. capricolum* tRNA genes, the sequence of both of which resembles that of $tRNA^{Trp}$ of *E. coli*. The anticodon sequence of one of these tRNA genes is 5′-T-C-A-3′, which can base-pair with both opal codon UGA and universal tryptophan codon UGG. That of the other is 5′-C-C-A-3′, which may base-pair exclusively with UGG. These two tRNA genes are expressed in the cell. All these findings suggest strongly that, in *M. capricolum*, UGA codes for tryptophan using the opal $tRNA_{UCA}$ but not $tRNA_{CCA}$.

## RESULTS AND DISCUSSION

**UGA Codons in *M. capricolum* S3 and L16 Genes.** As reported in a previous paper (1), we isolated the recombinant plasmid pMCB1088 containing a 9-kilobase-pair fragment of *M. capricolum* DNA. The fragment contains the genes for at least nine ribosomal proteins—S3, S5, S8, S14, S17, L5, L6, L16, and L18—as deduced from its encoded protein sequences being highly homologous with the corresponding *E. coli* ribosomal protein sequences (refs. 1 and 2; unpublished results). Fig. 1 shows the complete nucleotide sequence of a 629-base-pair (bp) *Hin*dIII fragment which is a part of the insert of pMCB1088 (see refs. 1 and 2). The DNA corresponds to the 3′ half of the S3 gene and about 90% of the L16 gene from the 5′ terminus. When the *M. capricolum* sequences are aligned with the *E. coli* protein sequences (3, 4) (Fig. 1), four UGA (opal) codons are found within the reading frames. The possibility that these UGA codons are termination signals can be excluded by their occurrence in the regions having extensive sequence homologies with the *E. coli* proteins. More importantly, three out of the four UGA codons appear at the positions corresponding to tryptophan in the *E. coli* proteins. This suggests that UGA is a sense codon, probably for tryptophan, in *M. capricolum*. No UGG codon for tryptophan has so far been found. One UGA appears in S3 at a site corresponding to arginine in the *E. coli* L16.

**Genes for Tryptophan tRNAs.** Since UGA, which is a stop codon in the universal code, seems to be read as tryptophan in *M. capricolum*, one would expect the occurrence of an opal tRNA that can decode the UGA codon. Plasmid pMCH964, having a 2.0-kilobase-pair *Hin*dIII fragment in pBR322, was isolated as one of the clones that hybridize with unfractionated *M. capricolum* tRNAs. By restriction mapping of this fragment followed by hybridization with $^{32}P$-labeled total *M. capricolum* tRNAs, the tRNA genes were localized within a 600-bp *Alu* I subfragment that had been derived from the middle part of the *Hin*dIII fragment (data not shown). The DNA sequence of this region (Fig. 2) revealed the presence of a pair of tRNA genes with a 40-bp spacer between them. The tRNA encoded by the first gene has an anticodon sequence 5′-U-C-A-3′ that can decode both opal codon UGA and universal tryptophan codon UGG, whereas the second one has an anticodon sequence 5′-C-C-A-3′ for the universal UGG codon for tryptophan (Fig. 3). The structural gene region for these two tRNAs is preceded by the expected promoter structures: a Pribnow-box-like sequence (underlined in Fig. 2) ≈20 bp upstream and a −35 sequence (also underlined in Fig. 2) 45 bp upstream from the coding sequences. The tRNA genes are followed by a probable termination signal: a dyad symmetrical structure and a stretch of thymidine residues (indicated by two arrows and by a broken line, respectively, in Fig. 2) 24 bp downstream from the coding sequence for $tRNA_{CCA}$. The above structure suggests that the two tRNA genes are arranged in a single operon. The $tRNA_{UCA}$ gene could have emerged by duplication of the $tRNA_{CCA}$ gene, since the two tRNA genes are closely related to each other not only in their tandem linkage on the chromosome but also in their high sequence homology (78% identity) and both $tRNA_{UCA}$ and $tRNA_{CCA}$ can be charged with tryptophan *in vitro* (see below).

**Expression of $tRNA^{Trp}$ Genes.** To determine whether the two tRNA genes described above are expressed *in vivo*, we purified tRNAs that hybridize with the DNA fragment containing these two tRNA genes. The purification procedure consists of hybridization of crude tRNAs with the DNA fragment, followed by identification of the hybridized tRNAs by sequencing. Since $tRNA_{UCA}$ is one base longer than $tRNA_{CCA}$, as deduced from their DNA sequences, they may

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviation: bp, base pair(s).

*On leave from the Department of Biochemistry and Biophysics, Research Institute for Nuclear Medicine and Biology, Hiroshima University, Hiroshima 734, Japan.
†Present address: Meiji Institute of Health Science, 540 Naruda, Odawara 250, Japan.

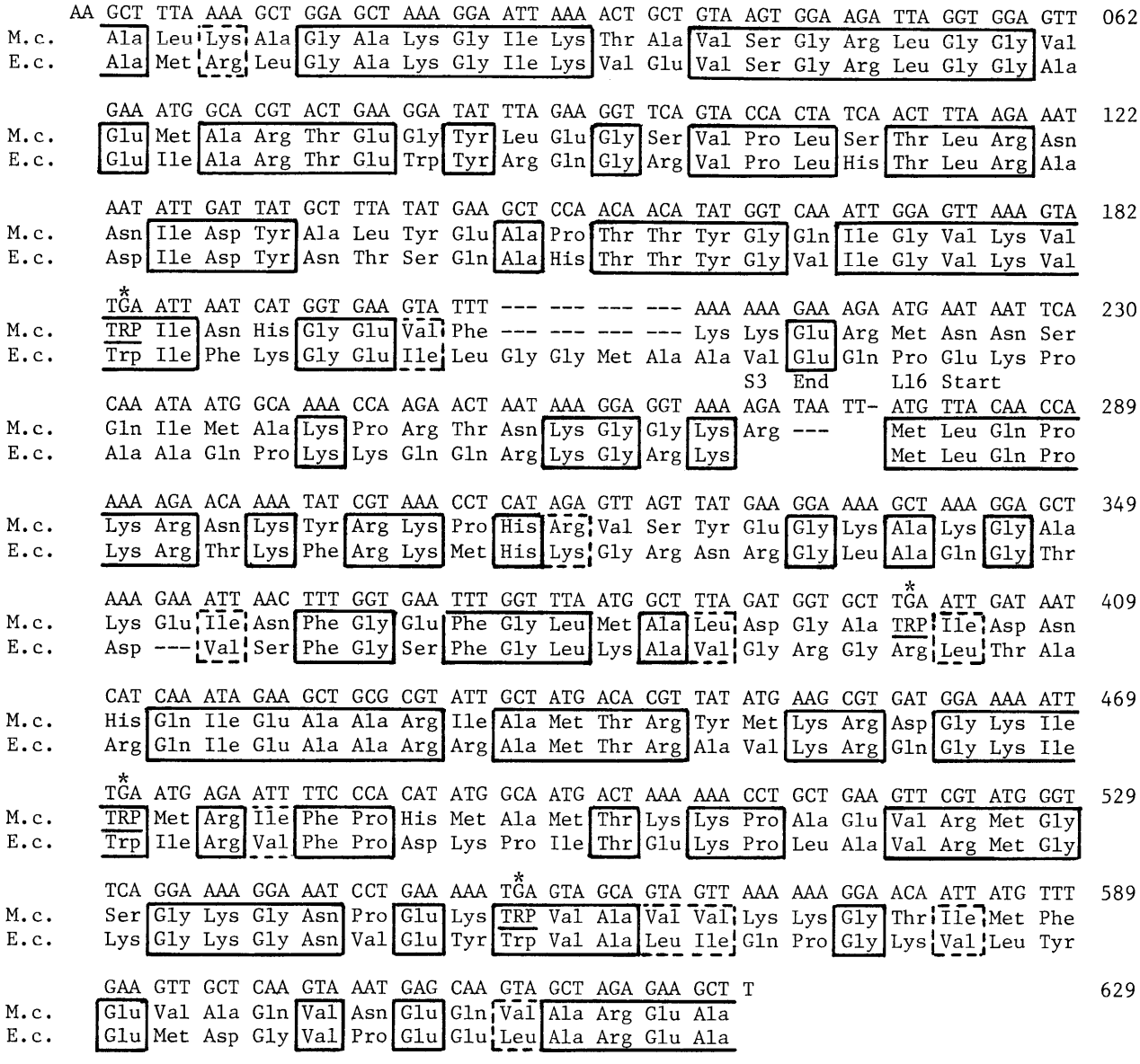


FIG. 1. DNA sequence of a part of *M. capricolum* S3 and L16 ribosomal protein genes. The DNA sequence of the mRNA-like strand, together with the predicted amino acid sequence, of a *Hin*dIII fragment from a plasmid pMCB1088 (see ref. 1) that contains a part of *M. capricolum* [American Type Culture Collection 27343 (Kid)] ribosomal protein S3 and L16 genes (M.c.) was aligned with the corresponding *E. coli* protein sequences (E.c.). The sites for identical amino acids and those of conservative amino acid substitution are boxed with solid and dotted lines, respectively. TGA triplets are marked with asterisks. DNA sequencing was performed by the chain-termination methods (5, 6).

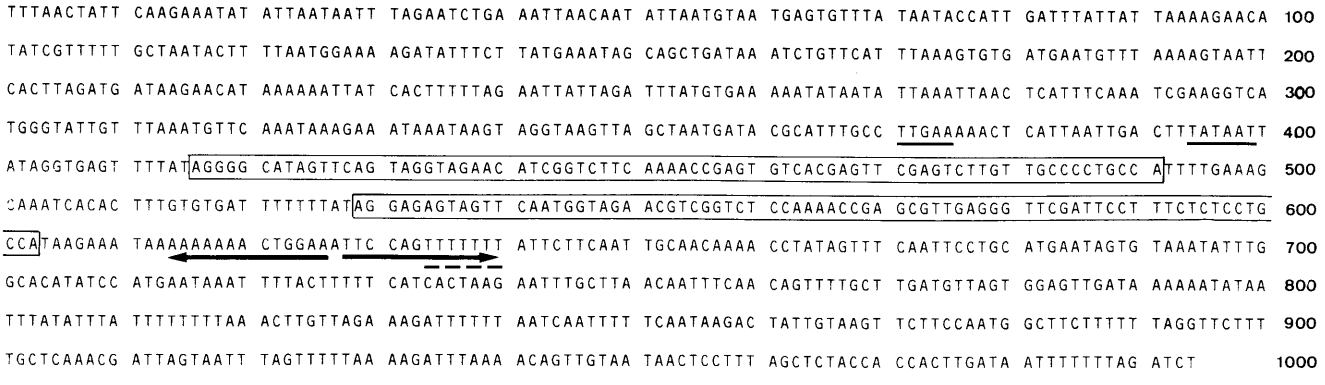


FIG. 2. DNA sequence of two tryptophan tRNA genes and their flanking region. Coding sequences for $tRNA_{UCA}$ and $tRNA_{CCA}$ are boxed. A Pribnow box-like structure and −35 sequences are underlined. Probable transcription-termination (dyad symmetrical) structure is shown by two arrows. A stretch of thymidine residues is shown by a broken line.

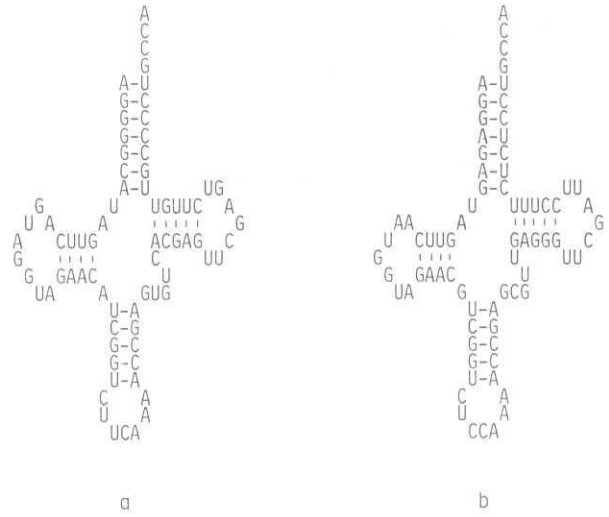



FIG. 3. Cloverleaf structures of $tRNA_{UCA}$ (*a*) and $tRNA_{CCA}$ (*b*) deduced from DNA sequence.

be easily distinguished by polyacrylamide gel electrophoresis. Fig. 4 shows an electrophoretic separation of tRNAs, the 3′ ends of which were labeled with [$^{32}$P]pCp and T4 RNA ligase after hybridization with the DNA fragment (see lane a). Only two tRNA species, differing slightly in length, predominated. Partial sequence analysis of these tRNAs by the chemical degradation method revealed that the fast migrating one agreed with the DNA sequence for the $tRNA_{CCA}$ gene, and the other, with that for the $tRNA_{UCA}$ gene (data not shown). Thus, in the cells, both the genes for $tRNA_{UCA}$ and $tRNA_{CCA}$ are transcribed and processed. However, when tRNAs were purified without first removing amino acids, $tRNA_{CCA}$ was predominantly labeled with [$^{32}$P]pCp (Fig. 4, lane b). This suggests that the bulk of the $tRNA_{UCA}$ molecules but not $tRNA_{CCA}$ have been already charged with amino acids *in vivo*, because amino acid bound to the -C-C-A terminus would prevent the ligation of [$^{32}$P]pCp to the 3′ end. To verify this and identify the amino acid bound to $tRNA_{UCA}$, the following experiment was performed. The deacylated tRNAs were incubated for a short period with amino acid in the presence of the *M. capricolum* S-100 fraction to reacylate the tRNAs. Charging the tRNAs with tryptophan inhibited the incorporation of [$^{32}$P]pCp into both $tRNA_{CCA}$ and $tRNA_{UCA}$ (Fig. 4, lane d), whereas other amino acids, leucine for example, did not affect the labeling efficiency (Fig. 4, lane c). These results indicate that both $tRNA_{CCA}$ and $tRNA_{UCA}$ accept tryptophan *in vitro*.

The presence of $tRNA^{Trp}_{UCA}$ in *M. capricolum* strongly supports the idea that UGA is translated as tryptophan by using this tRNA. Tryptophan is "universally" coded for by a single codon, UGG, which is decoded by $tRNA_{CCA}$ throughout prokaryotes (9) and eukaryotes (10). In mitochondria, not only UGG but also UGA are used as tryptophan codons (11–13), both of which are translated by a single tRNA with the anticodon UCA (14–17). Thus, the discovery of two tRNA species having anticodon sequences of, respectively, CCA and UCA in one genetic system contrasts with previous observations in other systems. Although the UCA anticodon can decode both UGA and UGG according to the wobble theory (18), we have not so far found UGG codons in the reading frames not only in the S3 and L16 genes but in other ribosomal protein genes (see also ref. 2). This suggests that UGA is predominantly, if not exclusively, used as a tryptophan codon in *M. capricolum*. It is thus interesting to see whether and how $tRNA_{CCA}$ participates in translation. The failure to find $tRNA_{CCA}$ appreciably charged *in vivo* with tryptophan might mean that its role is subsidiary.

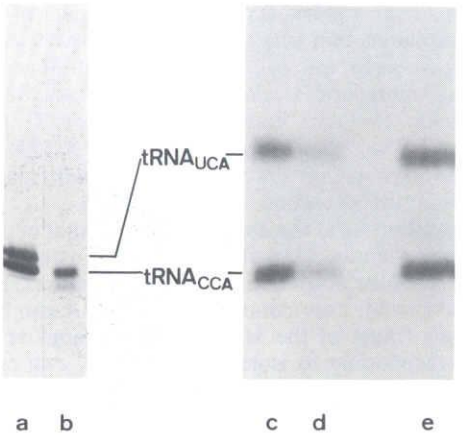



FIG. 4. Hybridization of tRNAs with pMCH964 DNA. Total tRNAs were prepared from *M. capricolum* by direct phenol extraction, followed by hybridization with pMCH964 DNA. For hybridization, pMCH964 DNA (1 mg) was fragmented by sonication and bound to a Sephacryl S-500 column according to the method of Bünemann and Westhoff (7). Hybridization was carried out at 42°C for 24 hr in a hybridization buffer containing 20 mM Pipes [piperazine-*N*,*N*′-bis(2-ethanesulfonic acid)] (pH 6.5), 0.6 M NaCl, 1 mM EDTA, 0.2% $NaDodSO_4$, and 50% (vol/vol) formamide. After unbound material was removed by washing with the hybridization buffer at 42°C, the hybridized tRNAs were eluted with the buffer at 90°C, precipitated, and washed with 70% (vol/vol) ethanol. The tRNAs so obtained were incubated at 37°C in 100 mM Tris·HCl (pH 10.0) for 30 min to "strip" amino acid (deacylate), if necessary. The 3′ ends of the tRNAs were labeled with [$^{32}$P]pCp and T4 RNA ligase as described by Peattie (8), followed by electrophoresis in 12% polyacrylamide gel and autoradiography. Electrophoresis was at 50 V/cm for 5 hr (lanes a and b) or at 10 V/cm for 30 hr (lanes c–e). Lanes a and b: tRNAs hybridized to pMCH964 DNA and labeled with [$^{32}$P]pCp after deacylation (a) or without deacylation (b). Lanes c and d: tRNAs were aminoacylated with leucine (c) or tryptophan (d) before [$^{32}$P]pCp labeling. Lane e: tRNAs were incubated with the S-100 fraction from *M. capricolum* without amino acid. Aminoacylation was carried out at 37°C for 5 min in a reaction mixture consisting of 50 μl of 0.2 M sodium cacodylate buffer (pH 7.0) containing 10 mM magnesium acetate and 8 mM ATP, 10 μl of amino acid solution (1 mM), 20 μl of the *M. capricolum* S-100 fraction, and 10 μl of $H_2O$ in a total volume of 100 μl.

**Evolutionary Aspects.** The A+T content of the *M. capricolum* genome is about 75%, one of the highest among all organisms. Reflecting this, an obvious preference for adenosine and thymidine in the *M. capricolum* genome has been seen in various regions: e.g., rRNA genes (19), their spacers (20), and codons (2). In a previous paper (2), we have demonstrated that the codons used for the *M. capricolum* ribosomal proteins S8 and L6 are strongly biased to those rich in A and U and that more than 90% of the codons have A or U at the third position. The same tendency of codon usage can be seen in the S3 and L16 genes, as shown in Fig. 1. The total A+T content of the coding regions is about 67%, and only 12 out of the total 207 codons have G or C at the third position [9 of these 12 are AUG (methionine) codons]. This suggests that the constraint for the preferential use of A and T in the protein genes is operating at the DNA level as a selection force. Thus, the use of UGA rather than UGG as a tryptophan codon in *M. capricolum* may be the consequence of this evolutionary pressure.

It has been suggested that mitochondria have evolved from certain prokaryotes by endosymbiosis (21, 22). The use of UGA as a tryptophan codon in *M. capricolum* as in mitochondria raises an interesting possibility in the phylogenetic relationship between mycoplasmas and mitochondria. Mycoplasmas are parasitic in eukaryotes, and the A+T-richness of their genomic DNAs resembles that of the mitochondrial

DNAs of lower eukaryotes. The codon usage of the yeast mitochondrial protein genes is strongly biased to the A- and U-rich codons (13, 23, 24), as is the case in *M. capricolum* (2). Furthermore, the *M. capricolum* $tRNA^{Trp}_{UCA}$ sequence is more similar to yeast mitochondrial $tRNA^{Trp}_{UCA}$ (14) (66% identity) than to cytoplasmic $tRNA^{Trp}_{CCA}$ (25) (55% identity). Thus, the mycoplasma-like organisms might have played some role in the evolution of mitochondria.

The deviation from the universal codons that occurs in mitochondria, such as AUA for methionine instead of isoleucine and AGA for nonsense instead of arginine (11), may not be the case in *M. capricolum*, because AUA can be seen in the reading frame of the *M. capricolum* genes at the positions corresponding to isoleucine in the *E. coli* protein sequences (e.g., the 47th codon of the L16 gene in Fig. 1) and AGA is the most abundantly used codon for arginine in this organism (2).

We thank Dr. Hiroshi Hori of this laboratory for many stimulating discussions. This work was supported by the Ministry of Education, Science, and Culture, Japan (Grants 59400007, 59107004, 59107010, and 59214010) and the Naito Science Foundation (Research Grant 81-106).

1. Kawauchi, Y., Muto, A., Yamao, F. & Osawa, S. (1984) *Mol. Gen. Genet.* **196,** 521–525.
2. Muto, A., Kawauchi, Y., Yamao, F. & Osawa, S. (1984) *Nucleic Acids Res.* **12,** 8209–8217.
3. Brauner, D. & Röming, R. (1979) *FEBS Lett.* **106,** 352–357.
4. Brosius, J. & Chen, R. (1976) *FEBS Lett.* **68,** 105–109.
5. Messing, J. & Vieira, J. (1982) *Gene* **19,** 269–276.
6. Sanger, F., Wicken, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74,** 5463–5467.
7. Bünemann, H. & Westhoff, P. (1983) *Methods Enzymol.* **100,** 400–407.
8. Peattie, D. A. (1979) *Proc. Natl. Acad. Sci. USA* **76,** 1760–1764.
9. Hirsh, D. (1971) *J. Mol. Biol.* **58,** 439–458.
10. Harada, F., Sawyer, R. C. & Dahlberg, J. E. (1975) *J. Biol. Chem.* **250,** 3487–3497.
11. Barrell, B. G., Bankier, A. T. & Drouin, J. (1979) *Nature (London)* **282,** 189–194.
12. Macino, G., Coruzzi, G., Nobrega, F. G., Li, M. & Tzagoloff, A. (1979) *Proc. Natl. Acad. Sci. USA* **76,** 3784–3785.
13. Fox, T. D. (1979) *Proc. Natl. Acad. Sci. USA* **76,** 6534–6538.
14. Martin, N. C., Pham, H. D., Underbrink-Lyon, K., Miller, O. L. & Donelson, J. E. (1980) *Nature (London)* **285,** 579–581.
15. Heckman, J. E., Sarnoff, J., Alzner-DeWeerd, B., Yin, S. & RajBhandary, U. L. (1980) *Proc. Natl. Acad. Sci. USA* **77,** 3159–3163.
16. Barrell, B. G., Anderson, S., Bankier, A. T., DeBruijn, M. H. L., Chen, E., Coulson, A. R., Drouin, J., Eperon, I. C., Nierlich, D. P., Roe, B. A., Sanger, F., Schreier, P. H., Smith, A. J. H., Staden, R. & Young, I. G. (1980) *Proc. Natl. Acad. Sci. USA* **77,** 3164–3166.
17. Köchel, H. G., Lazarus, C. M., Basak, N. & Küntzel, H. (1981) *Cell* **23,** 625–633.
18. Crick, F. H. C. (1966) *J. Mol. Biol.* **19,** 548–555.
19. Iwami, M., Muto, A., Yamao, F. & Osawa, S. (1984) *Mol. Gen. Genet.* **196,** 317–322.
20. Sawada, M., Muto, A., Iwami, M., Yamao, F. & Osawa, S. (1984) *Mol. Gen. Genet.* **196,** 311–316.
21. Margulis, L. (1981) *Symbiosis in Cell Evolution* (Freeman, San Francisco).
22. Küntzel, H. & Köchel, H. G. (1981) *Nature (London)* **293,** 751–755.
23. Dujon, B. (1980) *Cell* **20,** 185–197.
24. Macreadie, I. G., Novitski, C. E., Maxwell, R. J., John, U., Ooi, B., McMullen, G. L., Lukins, H. B., Linnane, A. W. & Nagley, P. (1983) *Nucleic Acids Res.* **11,** 4435–4451.
25. Keith, G., Roy, A., Ebel, J. P. & Dirheimer, G. (1972) *Biochimie* **54,** 1405–1426.

*Proc. Natl. Acad. Sci. USA*
Vol. 84, pp. 7398–7402, November 1987
Biochemistry

# Occurrence of unmodified adenine and uracil at the first position of anticodon in threonine tRNAs in *Mycoplasma capricolum*

(codon recognition/unmodified anticodon/biased mutation pressure)

YOSHIKI ANDACHI, FUMIAKI YAMAO, MASAFUMI IWAMI*, AKIRA MUTO, AND SYOZO OSAWA

Laboratory of Molecular Genetics, Department of Biology, Faculty of Science, Nagoya University, Chikusa-Ku, Nagoya 464, Japan

*Communicated by Motoo Kimura, June 18, 1987*

**ABSTRACT** **Codon usage pattern in the threonine four-codon (ACN) box in *Mycoplasma capricolum* is strongly biased towards adenine and uracil for the third base of codons. Codons ending in uracil or adenine, especially ACU, predominate over ACC and ACG. This bacterium contains two isoacceptor threonine tRNAs having anticodon sequences AGU and UGU, both with unmodified first nucleotides. It would thus appear that ACN codons are translated in an unusual way; $tRNA^{Thr}$(AGU) would translate the most abundantly used codon ACU exclusively, because adenine at the first anticodon position can, according to the wobble rule, pair only with uracil of the third codon position. The $tRNA^{Thr}$(UGU) would mainly be responsible for translation of three other codons, ACA, ACG, and ACC. Anticodon UGU would also be used for reading codon ACU as a redundancy of $tRNA^{Thr}$(AGU), as deduced from the mitochondrial code where unmodified uracil at the first anticodon position can pair with adenine, cytosine, guanine, and uracil by four-way wobble. The $tRNA^{Thr}$(AGU) has much higher sequence homology to $tRNA^{Thr}$(UGU) from *M. capricolum* (88%), *Bacillus subtilis* (77%) and *Escherichia coli* (86%) than to $tRNA^{Thr}$(GGU) from *B. subtilis* (66%) and *E. coli* (63%), suggesting that $tRNA^{Thr}$(AGU) has been derived from $tRNA^{Thr}$(UGU), but not from $tRNA^{Thr}$(GGU).**

In the eubacterial genetic code, the first bases of tRNA anticodons for four-codon boxes are generally guanine and uracil (uracil almost always modified), with, additionally, frequent occurrence of cytosine (see ref. 3). Anticodon GNN translates codons NNU and NNC, and anticodon hoUNN [hoU = a 5-hydroxyuridine derivative (see ref. 4)] translates codons NNA, NNG, and possibly NNU by wobbling (4). Anticodon CNN translates only NNG codons. Presumably, predominance of codons NNG in the G+C-rich bacterial lines leads to a need for anticodon CNN (3). The only exception is the arginine four-codon box (CGN), where anticodon ICG (I for the hypoxanthine of inosine) translates codons CGU, CGC, and CGA; and anticodon CCG translates codon CGG.

During analyses of the tRNA gene sequence in *Mycoplasma capricolum*, we have found a gene cluster containing five tRNA genes, one of which is a gene for putative threonine tRNA with anticodon sequence AGU. Anticodon AGU, unmodified, can, according to the wobble rule (1), translate only codon ACU, but not ACC. Because threonine codon ACC rarely appears in reading frames in *M. capricolum*, the most plausible explanation for the above finding would be that the first nucleoside of this tRNA anticodon is replaced by inosine (5) after transcription, so that anticodon IGU can translate codon ACC, as well as ACU and ACA, by three-way wobble, as in the case of ICG in the arginine four-codon box. Anticodon IGU appears in yeast (6). Another possibility would be that $tRNA^{Thr}$(GGU) exists in addition to $tRNA^{Thr}$(AGU), so that codons ACU and ACC can be translated, although coexistence of anticodons GNN and ANN (or INN) has never been reported.

We have thus analyzed the product of the gene for $tRNA^{Thr}$(AGU), as well as the other threonine isoacceptor tRNA(s). Unexpectedly, the results have shown that the two possibilities mentioned above are both incorrect; there exist two isoacceptor tRNAs, one having anticodon sequence AGU and the other having UGU and both having unmodified first nucleotides. Thus far, unmodified adenine at the first position of the anticodon has only been found for $tRNA^{Arg}$(ACG) in yeast mitochondria (7, 8).

## MATERIAL AND METHODS

**DNA Cloning and Sequencing.** The gene library of *Mycoplasma capricolum* [American Type Culture Collection 27343 (Kid.)] was constructed with *Hin*dIII-digested genomic DNA ligated to plasmid pBR322. The recombinant plasmid pMCH502 containing a 4.0-kilobase-pair (kbp) DNA insert was isolated as one of the clones that hybridized with unfractionated *M. capricolum* tRNA. DNA sequencing was done by the dideoxynucleotide chain-termination method (9).

**Isolation of tRNA by DNA Column.** Total tRNAs were prepared from *M. capricolum* by direct phenol extraction. A column of Sephacryl S-500 conjugated with 1.3-kbp *Acc* II fragment of the 4.0-kbp DNA (see above) was prepared according to the method of Bünemann and Westhoff (10). The 1.3-kbp fragment contained genes for $tRNA^{Thr}$(AGU) and a part of the 5′ end of $tRNA^{Tyr}$(GUA). Hybridization to and elution from the DNA column of the tRNAs were done as described (11). The eluted tRNAs were further purified by 12% polyacrylamide gel electrophoresis (11).

**Isolation of tRNA by Benzoylated DEAE-Cellulose Column.** Deacylated total tRNAs were incubated with threonine in the presence of the *M. capricolum* S100 fraction. The threonine-charged tRNAs were separated from other tRNAs by chromatography on benzoylated DEAE-cellulose column after 2-naphthoxyacetylation of the $tRNAs^{Thr}$ according to the method of Gillam *et al.* (12). The threonine tRNA fraction was then subjected to 12% gel electrophoresis to separate the tRNA species; the tRNA bands were eluted from the gel and used for RNA sequencing.

**RNA Sequencing.** The nucleotide sequence was determined by the chemical methods of Peattie (13) and the enzyme method of Donis-Keller (14) using 3′ or 5′ $^{32}P$-labeled RNA. The sequence was also determined by the postlabeling method of Kuchino *et al.* (15). Modified nucleotides were identified in the two-dimensional TLC using two different solvent systems: System I, isobutyric acid/0.5 M ammonia (5:3) in the first dimension and isopropyl alcohol/HCl/water (70:15:15) in the second dimension (16); System II, isobutyric

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

*Present address: Laboratory of Embryology, Department of Biology, Faculty of Science, Nagoya University.

acid/ammonia/water (60:1:33) in the first dimension and 100 ml of 0.1 M sodium sulfate (pH 6.8), 60 g of ammonium sulfate, and 2 ml of *n*-propylalcohol in the second dimension (17).

**Chemicals.** Enzymes were purchased from Takara-Shuzo (Kyoto), [5′-$^{32}$P]pCp and [$\gamma$-$^{32}$P]ATP were from Amersham (Japan), Sephacryl S-500 was from Pharmacia Fine Chemicals, benzoylated DEAE-cellulose was from Serva (Heidelberg), and cellulose thin-layer plates were from Funakoshi Pharmacological (Tokyo).

## RESULTS

We isolated the recombinant plasmid pMCH502 containing a 4-kbp *Hin*dIII fragment of *M. capricolum* DNA as one of the clones that hybridizes with unfractionated tRNA. The fragment contained a cluster of five tRNA genes. The order was 5′–tRNA$^{Thr}$(AGU)–tRNA$^{Tyr}$(GUA)–tRNA$^{Gln}$(UUG)–tRNA$^{Lys}$(UUU)–tRNA$^{Leu}$(UAA)–3′ (Fig. 1). The tRNA gene cluster was preceded by a putative promoter structure and followed by a probable termination signal. The tRNA genes were divided by short (4- to 8-base-pairs) spacers. Thus these genes seem to consist of a single transcriptional unit. Of special interest is that the anticodon sequence of tRNA$^{Thr}$ gene was AGU, because in the eubacterial code, the first base of anticodons is always guanine, uracil, or cytosine, with an exception of anticodon ICG for arginine (see above).

To determine whether the tRNA$^{Thr}$(AGU) gene described above is transcribed and the product tRNA has the anticodon sequence AGU or IAU, we purified the tRNAs that hybridized with a DNA fragment (a derivative of pMCH502) containing a complete sequence of tRNA$^{Thr}$(AGU) and the 5′-half of tRNA$^{Thr}$(GUA) genes. The purification procedure consists of hybridization of total tRNAs with the DNA fragment, followed by identification of the hybridized tRNAs by sequencing. Two tRNAs were detected; one of them agreed with the DNA sequence for the tRNA$^{Thr}$, and the other agreed with the DNA sequence for the tRNA$^{Tyr}$, as expected. The anticodon sequence of tRNA$^{Thr}$ was AGU in which adenine was unmodified as judged by TLC (see below).

In the eubacterial code, threonine codons ACU and ACC are read by anticodon GGU; and codons ACA, ACG, and possibly ACU are read by anticodon hoUCU. Anticodon AGU would be able to recognize only ACU, but not ACC. Table 1 shows that threonine codon usage pattern in *M. capricolum* is strongly A+T(or U)-biased, as expected from a high A+T content of the genome (18). Codons ending in uracil or adenine, especially ACU, predominated, and yet codon ACC was used in a low frequency. Codon ACG has not been found thus far.

Table 1. Threonine codon usage in *Mycoplasma capricolum*

| Codons | ACU | ACC | ACA | ACG |
|---|---|---|---|---|
| Occurrence* | 131 | 4 | 85 | 0 |
| Percent | (59.5) | (1.8) | (38.6) | (0) |

*Number of threonine codons in 20 ribosomal protein genes, *secY* and *adk* of *M. capricolum* (from ref. 27).

We then isolated the isoacceptor threonine tRNAs, and their sequences were determined. The deacylated total tRNAs were incubated for a short period with threonine in the presence of the *M. capricolum* S100 fraction to reacylate threonine tRNAs. The charged tRNAs were 2-naphthoxyacetylated and separated from other nonacylated tRNAs on a benzoylated DEAE-cellulose column, followed by further purification with polyacrylamide gel electrophoresis. Fig. 2 shows an electrophoretic separation of the tRNAs. Three RNAs (a, b, and c in Fig. 2), differing slightly in length, predominated. A longer RNA (d) was also found as a minor component. All the four RNA species were eluted and sequenced. The band d was tRNA$^{Leu}$(UAA). The tRNAs of band a and b were tRNA$^{Thr}$(AGU) and tRNA$^{Thr}$(UGU), respectively. The tRNA of band c had anticodon sequence NAU (N, unidentified modified base) and is probably tRNA$^{Ile}$. The first modified base (*C) of the anticodon for the *E. coli* tRNA$^{Ile}$(*CAU) contains a nonaromatic amino acid (Yokoyama *et al.*, personal communication). If the same is applied to the *M. capricolum* tRNA$^{Ile}$(NAU), then the amino acid residue in this tRNA would be 2-naphthoxyacetylated and be recovered together with tRNA$^{Thr}$. TLC identification, using two different solvent systems, of the first (5′) nucleotide of anticodons clearly revealed that adenine in tRNA$^{Thr}$(AGU) and uracil in tRNA$^{Thr}$(UGU) were both unmodified; radioactive spots of these nucleotides perfectly agreed with the UV spots of authentic adenylic acid and uridylic acid, respectively (Fig. 3). Fig. 4 shows the total sequences of the two threonine tRNAs including modified nucleotides.

Because the bulk, if not all, of the threonine-charged tRNAs should be recovered from the benzoylated DEAE-cellulose chromatography, we conclude that *M. capricolum* contains two isoacceptor tRNAs with anticodon sequences of

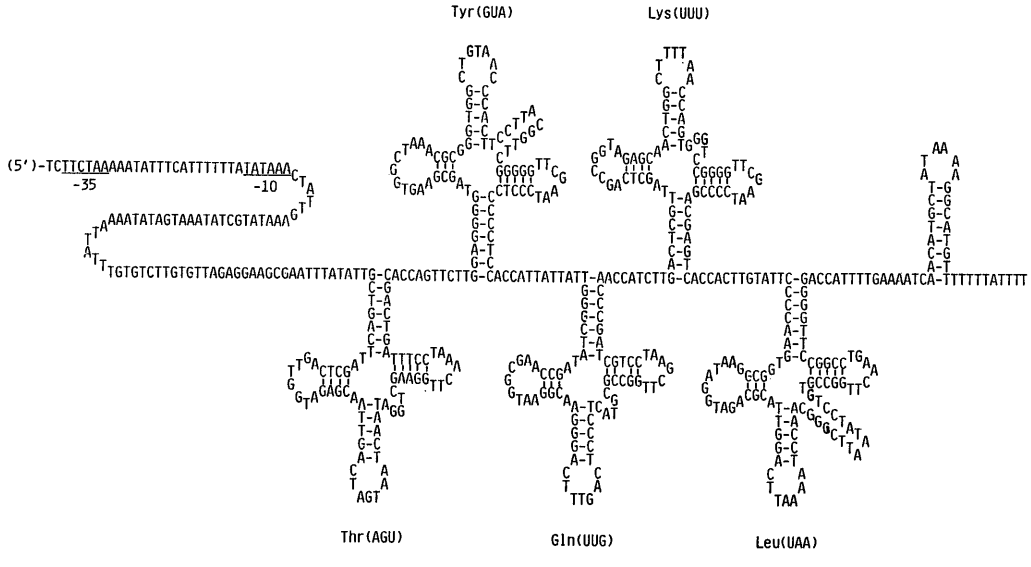



FIG. 1. DNA sequence of a *M. capricolum* tRNA gene cluster. Noncoding (RNA-like) strand of a part of pMCB502 DNA including tRNA genes is shown.


FIG. 2. Polyacrylamide gel electrophoresis of tRNAs. Deacylated total tRNAs of *M. capricolum* were incubated for 10 min at 37°C with 2 mM threonine in the presence of the S100 fraction to acylate threonine tRNAs. The charged tRNAs were 2-naphthoxyacetylated and separated from noncharged tRNAs by a benzoylated DEAE-cellulose column. (Lane 1) Total tRNAs; (lane 2) "charged" tRNAs isolated by benzoylated DEAE-cellulose column chromatography; (lane 3) $tRNA^{Thr}$(AGU) isolated by using a column of Sephacryl S-500 conjugated with the DNA fragment containing the gene for $tRNA^{Thr}$(AGU). a, $tRNA^{Thr}$(AGU); b, $tRNA^{Thr}$(UGU); c, (NAU; N, unidentified modified base) probably $tRNA^{Ile}$; and d, $tRNA^{Leu}$(UAA).

AGU and UGU. The first anticodon nucleotide is not modified in either of these tRNAs.

## DISCUSSION

We have shown that in *M. capricolum* there exist at least two threonine isoacceptor tRNAs having anticodon sequences of AGU and UGU, the first bases adenine and uracil being both unmodified. No tRNA having anticodon GGU or CGU has been found, although a possibility of the presence of these two tRNA species as minor components cannot be excluded. Assuming that $tRNA^{Thr}$(AGU) and $tRNA^{Thr}$(UGU) are the only threonine tRNAs in this organism, codon ACU, which is the most abundantly used threonine codon (18) (see Table 1), would appear to be translated by anticodon AGU. Anticodon UGU would mainly be responsible for three other threonine codons, ACA, ACG, and ACC. Because unmodified uracil at the first anticodon position has been reported to be able to pair with adenine, guanine, cytosine, and uracil by four-way wobble in the mitochondrial code (2), $tRNA^{Thr}$(UGU) could also be used for reading codon ACU as a redundancy of $tRNA^{Thr}$(AGU).

As postulated previously (3), the early code would have used GNN anticodon for translation of codons NNY (where Y = uracil or cytosine) and anticodon hoUNN for codons NNR (where R = adenine or guanine) in all the four-codon boxes. This rule can be applied to most of the eubacterial code, except for the arginine four-codon box. The evolution of AGU and UGU anticodons in the *Mycoplasma* line can be deduced as follows.

The simplest explanation for the appearance of anticodon AGU would be that GGU mutated to AGU by A+T-biased mutation pressure as assumed from a very high A+T-content (75%) of the *M. capricolum* genomic DNA (19). However, this seems not to be the case. The $tRNA^{Thr}$(AGU) would most probably have been derived from $tRNA^{Thr}$(UGU), but

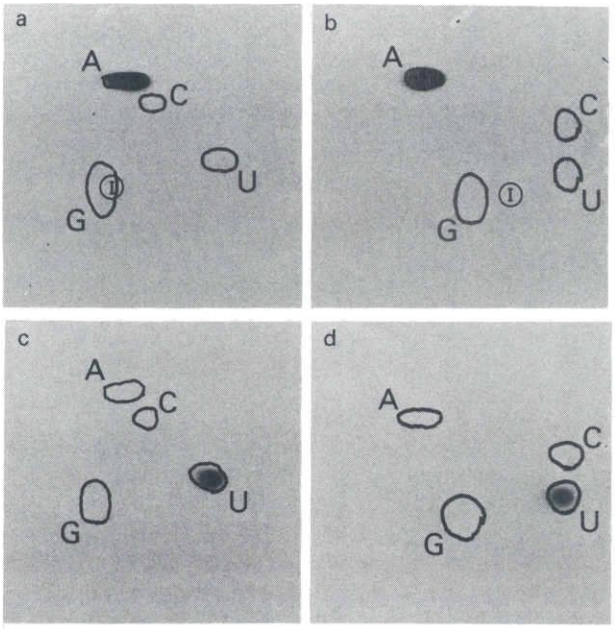



FIG. 3. TLC identification of the first (5′) nucleotide of anticodons of $tRNA^{Thr}$(AGU) and $tRNA^{Thr}$(UGU). $tRNA^{Thr}$(AGU) and $tRNA^{Thr}$(UGU) were isolated by DNA column and benzoylated DEAE-cellulose column, respectively. The tRNA was partially digested with formamide, and the 5′ ends of the digests were labeled with [$\gamma$-$^{32}$P]ATP and polynucleotide kinase. The labeled products were separated by 20% polyacrylamide gel electrophoresis according to their chain lengths and autoradiographed. The 5′ end-labeled oligonucleotide in each radioactive band was eluted from the gel and digested with ribonuclease P1 to produce N-[5′-$^{32}$P]monophosphate. The product was mixed with nonlabeled specific N,N,N,N-5′-monophosphates and displayed by two-dimensional TLC (15). The anticodon sequence could be identified from the total tRNA sequence determined by the above method. To see the modification of the first nucleotide of anticodon, the corresponding nucleotide was further identified by TLC. The labeled first nucleotide of the anticodon from $tRNA^{Thr}$(AGU) (*a* and *b*) or from $tRNA^{Thr}$(UGU) (*c* and *d*) was displayed by TLC using solvent system I (*a* and *c*) and system II (*b* and *d*), respectively, and exposed on an x-ray film. Positions of authentic unmodified nucleotides were determined by UV irradiation. I (inosinic acid) should be positioned as indicated in the figures. Positions of various modified uridine nucleotides do not agree with that of the unmodified one (see refs. 16 and 17).

not from $tRNA^{Thr}$(GGU), as assumed from the following facts. The *M. capricolum* tRNA gene arrangement, $tRNA^{Thr}$(AGU)–$tRNA^{Tyr}$(GUA)–$tRNA^{Gln}$(UUG)–$tRNA^{Lys}$-(UUU)–$tRNA^{Leu}$(UAA) studied here (Fig. 1), is similar to a part of the *B. subtilis* "16 tRNA gene cluster," the arrangement of which is *$tRNA^{Thr}$(UGU)–$tRNA^{Tyr}$(GUA)*–$tRNA^{Trp}$-(CCA)–$tRNA^{His}$(GUG)–*$tRNA^{Gln}$(UUG)*–$tRNA^{Gly}$(GCC)–$tRNA^{Cys}$(GCA)–*$tRNA^{Leu}$(UAA)* (20). The *B. subtilis* genes shown in italic letters are those also found in *M. capricolum*. The sequence similarity between presumed homologous tRNA genes of *M. capricolum* with those of *B. subtilis* is high—i.e., 79% between $tRNA^{Thr}$(AGU) (*M, M. capricolum*) and $tRNA^{Thr}$(UGU) (*B, B. subtilis*), 78% between $tRNA^{Tyr}$(GUA) (*M*) and $tRNA^{Tyr}$(GUA) (*B*), 90% between $tRNA^{Gln}$(UUG) (*M*) and $tRNA^{Gln}$(UUG) (*B*), and 69% between $tRNA^{Leu}$(UAA) (*M*) and $tRNA^{Leu}$(UAA) (*B*). On the other hand, *B. subtilis* $tRNA^{Thr}$(GGU), the gene of which is a member of the "6 tRNA gene cluster" (21), has only 64% similarity to the *M. capricolum* $tRNA^{Thr}$(AGU). These observations suggest that the *M. capricolum* $tRNA^{Thr}$(AGU) gene has resulted from a mutation of the $tRNA^{Thr}$(UGU) gene that is homologous with the *B. subtilis* $tRNA^{Thr}$(UGU) gene mentioned above. *B. subtilis* has one more tRNA gene cluster—i.e., the "21 tRNA gene cluster,"

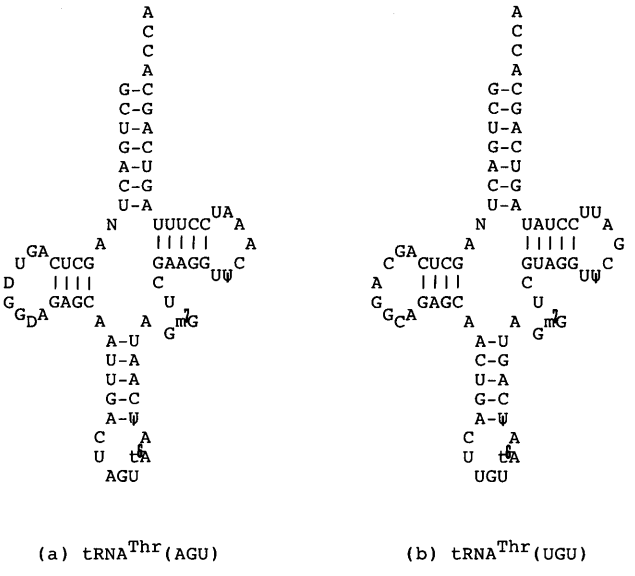



FIG. 4. Structures of the two threonine tRNAs of *M. capricolum*. The two tRNAs were purified through gel electrophoresis (see Fig. 2), and sequence was determined by chemical (13) and enzyme method (14) and also by a postlabeling method (15). Modified nucleotides were identified by TLC (16, 17). N represents unidentified modified uracil. (*a*) tRNA$^{Thr}$(AGU) and (*b*) tRNA$^{Thr}$(UGU).

containing the second tRNA$^{Thr}$(UGU) gene; its gene arrangement is unrelated to the "16 tRNA gene cluster," and yet the sequence of the second tRNA$^{Thr}$(UGU) is identical with that (first) in the "16 tRNA gene cluster" (20, 22). The sequence similarities of the first and the second *B. subtilis* tRNA$^{Thr}$-(UGU) to *M. capricolum* tRNA$^{Thr}$(UGU) are high (77%), thus suggesting that the second *B. subtilis* tRNA$^{Thr}$(UGU) gene in the "21 tRNA gene cluster" and *M. capricolum* tRNA$^{Thr}$(UGU) gene are homologous counterparts. *M. capricolum* tRNA$^{Thr}$(AGU) also reveals a higher similarity to *E. coli* tRNA$^{Thr}$(UGU) (86%) than to tRNA$^{Thr}$(GGU) (63%).

From these observations, two possibilities would exist for the evolution of threonine tRNAs in the *Mycoplasma* line. (*i*) First, anticodon UGU became unmodified as a result of partial deprivation of the uracil-modification enzyme system, so that UGU could translate all the threonine codons by four-way wobble as deduced from the mitochondrial code, and GGU disappeared. This process would be simultaneous with a tendency to genomic economization in the *Mycoplasma* line (23). An increasingly A+T-biased mutation pressure led to an extreme predominance of ACU (and ACA) codons over ACC and ACG in the genes. Under these circumstances, one of the genes for tRNA$^{Thr}$(UGU) mutated to tRNA$^{Thr}$-(AGU). (*ii*) UGU became unmodified, followed by mutation of one of the genes for UGU to AGU. This would be a transient stage where UGU, AGU, and GGU all existed. GGU then disappeared, for the low demand for translation of codon ACC would be supplied by anticodon UGU, and anticodon GGU was not needed. In any case, the appearance of tRNA$^{Thr}$(AGU) is adaptive to fulfill a heavy demand for translation of codon ACU. This situation may be somewhat analogous to an increase in G+C-rich bacterial lines in the content of anticodons CNN, which would have been called for by a predominance of NNG codons in the protein genes (see *Introduction*).

One significance of the presence of anticodon AGU would be that, as mentioned above, it would translate codon ACU more efficiently than anticodon UGU. Moreover, the codon–anticodon pairing 5′ACN3′·3′UGU5′ involves only one G·C pair, which might cause misreading. Thus the presence of anticodon AGU would be advantageous in view of both efficiency and correct reading of the most frequently used ACU threonine codon in this organism. In this connection, it is of interest to note that *Mycoplasma mycoides* seems to contain only one species of glycine tRNA having anticodon sequence UCC (24, 25). Because codon–anticodon pairing 5′GGN3′·3′CCU5′ in glycine four-codon box involves two G·C pairs, all four codons can be translated by four-way wobble more correctly, as compared with the case of the threonine box. The four-codon boxes where two G·C pairs are involved in codon recognition by the second and the third positions of the anticodon are for proline, alanine, and arginine in addition to glycine, whereas those containing only one G·C pair are for leucine, valine, and serine in addition to threonine. It is thus possible that codons in the former boxes containing one G·C pair could be read by anticodons ANN and UNN (N = cytosine or guanine, adenine or uracil), whereas in the latter containing two G·C pairs, four codons could be read only by anticodon UNN (N = guanine or cytosine), with an exception of arginine four-codon box. As already pointed out, all the eubacterial species so far studied use anticodon ICG for translation of arginine codons CGU, CGC, and CGA and anticodon CCG for codon CGG. In *M. mycoides*, a gene for tRNA with anticodon ACG has been reported (26), although whether adenosine is replaced by inosine after transcription and whether anticodon CCG exists have yet to be studied.

We thank Drs. Yoshiro Shimura, Thomas H. Jukes, Shigeyuki Yokoyama, and Yoshiyuki Kuchino for advice and comments. This work was supported by grants from the Ministry of Education, Science and Culture of Japan and Naito Science Foundation Research Grants.

1. Crick, F. H. C. (1966) *J. Mol. Biol.* **19,** 548–555.
2. Barell, B. G., Anderson, S., Bankier, A. T., DeBuijn, M. H. L., Chen, E., Coulson, A. R., Drouin, J., Eperon, I. C., Nierlich, D. P., Roe, B. A., Sanger, F., Schreier, P. H., Smith, A. J. H., Staden, R. & Young, I. G. (1980) *Proc. Natl. Acad. Sci. USA* **77,** 3164–3166.
3. Osawa, S., Jukes, T. H., Muto, A., Yamao, F., Ohama, T. & Andachi, Y. (1987) *Cold Spring Harbor Symp. Quant. Biol.* **52,** in press.
4. Yokoyama, S., Watanabe, T., Murao, K., Ishikura, H., Yamaizumi, Z., Nishimura, S. & Miyazawa, T. (1985) *Proc. Natl. Acad. Sci. USA* **82,** 4905–4909.
5. Elliott, M. S. & Trewyn, R. W. (1984) *J. Biol. Chem.* **259,** 2407–2409.
6. Wessenbach, J., Kiraly, I. & Dirheimer, G. (1977) *Biochimie* **59,** 381–388.
7. Lewin, B. (1985) *Genes II* (Wiley, New York), p. 127.
8. Sibler, A.-P., Dirheimer, G. & Martin, R. P. (1986) *FEBS Lett.* **194,** 131–138.
9. Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74,** 5463–5467.
10. Bünemann, H. & Westhoff, P. (1983) *Methods Enzymol.* **100,** 400–407.
11. Yamao, F., Muto, A., Kawauchi, Y., Iwami, M., Iwagami, S., Azumi, Y. & Osawa, S. (1985) *Proc. Natl. Acad. Sci. USA* **82,** 2306–2309.
12. Gillam, I., Blew, D., Warrington, R. C., von Tigerstrom, M. & Tener, G. M. (1968) *Biochemistry* **7,** 3459–3468.
13. Peattie, D. A. (1979) *Proc. Natl. Acad. Sci. USA* **76,** 1760–1764.
14. Donis-Keller, H. (1979) *Nucleic Acids Res.* **8,** 3133–3143.
15. Kuchino, Y., Kato, M., Sugisaki, H. & Nishimura, S. (1979) *Nucleic Acids Res.* **6,** 3459–3469.
16. Nishimura, S. (1979) in *Transfer RNA: Structure, Properties, and Recognition*, eds. Schimmel, P. R., Söll, D. & Abelson, J. M. (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY), pp. 59–79.
17. Silberklang, M., Prochiantz, A., Haenni, A. L. & RajBhandary, D. L. (1977) *Eur. J. Biochem.* **72,** 465–478.

18. Muto, A., Kawauchi, Y., Yamao, F. & Osawa, S. (1984) *Nucleic Acids Res.* **12,** 8209–8217.
19. Muto, A. & Osawa, S. (1987) *Proc. Natl. Acad. Sci. USA* **84,** 166–169.
20. Wawrousek, E. F., Narasimhan, N. & Hansen, J. N. (1984) *J. Biol. Chem.* **259,** 3694–3702.
21. Wawrousek, E. F. & Hansen, J. N. (1983) *J. Biol. Chem.* **258,** 291–298.
22. Green, C. J. & Vold, B. S. (1983) *Nucleic Acids Res.* **11,** 5763–5774.
23. Kawauchi, Y., Muto, A. & Osawa, S. (1982) *Mol. Gen. Genet.* **188,** 7–11.
24. Kilpatrick, M. & Walker, R. T. (1980) *Nucleic Acids Res.* **8,** 2783–2786.
25. Samuelsson, T., Axberg, T., Boren, T. & Langerkvist, U. (1983) *J. Biol. Chem.* **258,** 13178–13184.
26. Samuelsson, T., Elias, P., Lustig, F. & Guindy, Y. S. (1985) *Biochem. J.* **232,** 223–228.
27. Ohkubo, S., Muto, A., Kawauchi, Y., Yamao, F. & Osawa, S. (1987) *Mol. Gen. Genet.*, in press.

JOURNAL OF CELLULAR PHYSIOLOGY 136:319-325 (1988)

# Phorbol Esters Can Persistently Replace Interleukin-2 (IL-2) for the Growth of a Human IL-2-Dependent T-Cell Line

TOSHIKAZU TAKESHITA, YUSO GOTO, MASATAKA NAKAMURA, MASAHIRO FUJII, MASAFUMI IWAMI, YORIO HINUMA, AND KAZUO SUGAMURA*

*Department of Bacteriology, Tohoku University School of Medicine, Sendai 980 (T.T., Y.G., K.S.), Institute for Virus Research, Kyoto University, Kyoto 606 (M.N., M.F., Y.H.), and Molecular Oncology Division, Meiji Institute of Health Science, Odawara 250 (M.I.), Japan*

A selected clone from an IL-2-dependent human T-cell line was persistently propagated in the presence of phorbol esters with the ability to activate protein kinase C (PKC), such as 12-O-tetradecanoylphorbol-13-acetate (TPA) or phorbol-12,13-dibutylate (PDBu). Thus, a TPA(PDBu)-dependent T-cell line, designated TPA-Mat, was established from IL-2-dependent T cells. The TPA-dependency of TPA-Mat was not lost during cultivation for more than a year in the presence of TPA, and TPA-Mat cells still showed IL-2-dependent growth. However, the TPA (PDBu)-dependent growth of TPA-Mat did not seem to be mediated by an autocrine mechanism of IL-2 or by any other growth factor production, because these factors were not detected in TPA-Mat cell supernatants. Therefore, the phorbol esters substituted for IL-2 and may be directly involved in transduction of growth signals in TPA-Mat cells. Although activity of PKC was down-regulated, messenger ribonucleic acid (mRNA) of the PKC β-gene was detected in TPA-Mat cells cultured with PDBu. Furthermore, the growth of TPA-Mat cells was stimulated not only by phorbol esters but also by nonphorbol ester tumor promoters with the ability to activate PKC. These observations suggest that the sustained activation of PKC by the phorbol esters could induce continuous growth of the IL-2-dependent TPA-Mat cells.

In the immune response, resting T cells initially are activated by a specific antigen in the presence of antigen-presenting cells such as macrophages. The activation of T cells induces their production of IL-2 and expression of IL-2 receptor (IL-2R). Consequently, they are able to proliferate through an IL-2-dependent mechanism and to expand clonally. Thus, IL-2 and IL-2R are central to T cell proliferation. Recent studies demonstrated that intracellular growth signaling can be transduced from high affinity IL-2R after IL-2 binding (Robb et al., 1981; Smith and Cantrell, 1985; Depper et al., 1985). Receptors for several ligands—such as epidermal growth factor (EGF), insulin, platelet-derived growth factor (PDGF), and colony stimulating factor 1 (CSF-1)—are known to associate with tyrosine-kinase, which could have a crucial role in the triggering of signal transduction (Sibley et al., 1987; Yeung et al., 1987). The putative IL-2R (i.e., Tac molecule) is considered to have a cytoplasmic domain that is too small to harbor any kinase activity, although non-Tac molecule(s) in the high-affinity IL-2R have been identified recently (Sharon et al., 1986; Tsudo et al., 1986; Teshigawara et al., 1987; Robb et al., 1987). The molecular basis of these non-Tac molecule(s) has not yet been elucidated, and the mechanisms of signal transduction from the IL-2R remain to be resolved.

As IL-2R expression of activated T cells is transient in the normal immune response, spontaneous IL-2-dependent human T-cell lines rarely are established in vitro. However, in the murine system, IL-2-dependent T-cell lines often have been obtained spontaneously from in vitro culture with IL-2-containing medium. On the other hand, IL-2-dependent human T-cell lines have been established easily, but only when T cells are infected with human T-cell leukemia virus type I (HTLV-I). As HTLV-I infection results in the expression of IL-2R, the IL-2R expression is thought to correlate with HTLV-I-induced T-cell transformation (Sugamura et al., 1984; Inoue et al., 1986). Therefore, the signal transduction from IL-2R must be investigated for elucidation of regulatory mechanisms of malignant as well as normal T-cell growth.

Phorbol esters, such as TPA and PDBu, are potent tumor promoters in vivo (Berenblum, 1941; Hecker, 1971; Blumberg, 1980) and activators of PKC, a calcium- and phospholipid-dependent protein kinase (Castagna et al., 1982; Niedel et al., 1983; Nishizuka,

Received September 18, 1987; accepted March 22, 1988.

*To whom reprint requests/correspondence should be addressed.

M. Fujii is now at Molecular Biology and Virology Laboratory, The Salk Institute, P. O. Box 85800, San Diego, CA 92138.

1984). They induce transformed phenotypes (Blumberg et al., 1976; Weinstein et al., 1977; Weinstein and Wigler, 1977) and differentiation in culture (Huberman and Callaham, 1979; Lotem and Sachs, 1979) and are able to stimulate growth (Touraine et al., 1977; Rozengurt et al., 1984; Kaneshima et al., 1983) of various cell types in vitro. Transient growth of IL-2-dependent murine T-cell lines is induced by TPA (Kamber, 1986; Kim et al., 1986), and IL-2 induces rapid activation of PKC in an IL-2-dependent murine T-cell line in accordance with its growth promotion (Farrar and Anderson, 1985). These findings suggest that PKC is important in signal transduction from the IL-2R and that activation of PKC mediates the continuous growth of IL-2-dependent cells. In an effort to prove this, we established a TPA(PDBu)-dependent human T-cell line (TPA-Mat) from a HTLV-I-infected T-cell line that required IL-2 for growth.

## MATERIALS AND METHODS

### Cell culture

ILT-Mat cells were maintained in growth medium (RPMI 1640 medium supplemented with 10% fetal calf serum) containing 200 U/ml of human recombinant IL-2 (obtained from Shionogi Pharmaceutical Co., Osaka, Japan), and TPA-Mat cells were maintained in growth medium containing 81 nM TPA at 37°C with 7% $CO_2$ in air. The medium was renewed every 2 or 3 days by centrifugation.

### Incorporation of $^3$H-thymidine

In 200 μl of appropriate medium in 96-well microplates (Falcon No. 3072), $2.5 \times 10^4$ cells/well were incubated for 48 hr at 37°C under 7% $CO_2$ in air. During the last 4 hr of incubation, 1 μCi/well of $^3$H-thymidine ($^3$H-TdR) (Amersham) was added, and cells then were harvested. The incorporated $^3$H-TdR was measured in a liquid scintillation counter. Values are means for triplicate determinations.

### Southern and Northern blot hybridization

High-molecular-weight genomic DNA (15 μg) extracted from cells was digested with EcoRI, electrophoresed through a 0.7% agarose gel, and transferred to a nitrocellulose membrane (Schleicher & Schuell, Dassel, FRG) (Southern, 1975). The T-cell receptor β-chain constant region cDNA probe from 86T5 (Hedrick et al., 1984) and HTLV-I pX probe from pHTD2 (Ohtani et al., 1987) were labeled by nick translation and hybridized under 50% formamide at 42°C for 24 hr (Maniatis et al., 1982). The membranes were washed three times for 15 minutes each at 65°C in 20 × SSC (3 M NaCl and 0.25 M sodium citrate) containing 0.1% sodium dodecyl sulfate and exposed to X-ray films with intensifying screens at −70°C.

Total cellular RNA was isolated from cells by extraction with guanidine thiocyanate and centrifugation of extracts through a 5.7 M cesium chloride cushion as described previously (Glisin et al., 1974). Samples of total RNA (20 μg) were subjected to electrophoresis on a 1% agarose gel containing 2.2 M formaldehyde (Lehrach et al., 1977), transferred to a nitrocellulose membrane in 20 × SSC, and hybridized as described for Southern blots.

### Detection of PKC activity

Cells ($2.5 \times 10^7$) were harvested for preparation of subcellular fractions. Subcellular fractionation was carried out as reported by Farrar and Anderson (1985) with some modifications. The harvested cells were washed three times with phosphate-buffered saline (PBS) and once with 0.33 M sucrose by centrifugation. The cells were lysed in 1.0 ml of ice-cold sample buffer (20 mM Tris-HCl pH 7.5, 0.33 M sucrose, 2 mM EDTA, 0.5 mM EGTA, and 2 mM phenylmethylsulphonyl fluoride) by using a Terumo syringe 30 times. The cytosol and particulate membrane fractions of the lysates were prepared by centrifugation at 100,000 g for 60 minutes. The pelleted membrane fraction was solubilized in 1 ml of sample buffer containing 0.5% NP-40 by gentle homogenization with a glass rod for 30 minutes at 4°C and centrifuged at 12,000 rpm for 15 minutes and 1.0 ml of soluble membrane or cytosol preparation was applied to a 1.0 ml DE52 column pre-equilibrated with sample buffer minus sucrose in order to purify PKC, as described elsewhere (Kitano et al., 1986). Columns were washed with 15ml of sample buffer and then eluted with 2.0 ml of 0.08 M NaCl. Eluates were dialysed against 300 ml of buffer (20 mM Tris-HCl pH 7.5, 50 mM 2-mercaptoethanol and 60% glycerol), and dialysed aliquots (25 μl) were assayed for PKC activity in reaction mixtures (125 μl) containing 1 mM $CaCl_2$, 10 mM magnesium acetate, 100 μM [γ-$^{32}$P]ATP (60 cpm/pmol), and 25 μg of histone type V-S (Sigma) in the presence or absence of 2.5 μg of phosphatidylserine and 0.5 μg of diolein. The reaction was carried out for 10 minutes at 30°C and terminated by the addition of 1 ml of 0.1 sodium pyrophosphate containing 10 mM EDTA. The reaction mixtures were immediately passed through nitrocellulose membrane filters, and they were washed with 15 ml of 0.1 M sodium pyrophosphate containing 10 mM EDTA. The $^{32}$P radioactivity on the membrane filters was measured in a liquid scintillation counter. The indicated PKC activity was calculated by subtracting the incorporated radioactivity in the absence of phospholipids from that in the presence of phospholipids.

### Mono Q column chromatography

Cells ($2 \times 10^8$) were lysed in 6 ml of sample buffer containing 0.5% NP-40 and centrifuged at 12,000 rpm for 15 minutes. The supernatant was applied to a Mono Q HR5/5 column (Pharmacia) pre-equilibrated with sample buffer minus sucrose. The column was washed with 30 ml of sample buffer minus sucrose and then eluted with a linear gradient of NaCl (0.0–0.6 M) in sample buffer minus sucrose at a flow rate of 0.5 ml/minutes, and fractions of 1 ml were collected.

### $^3$H-PDBu binding

The $^3$H-PDBu binding assay was done in tubes using a reaction mixture (200 μl) consisting of 20 mM Tris-HCl (pH 7.5), 1 mM $CaCl_2$, 10 mM magnesium acetate, 4 μg of phosphatidylserine, 80 nM $^3$H-PDBu (15.8 Ci/mmol, New England Nuclear), and 25 μl of sample obtained from Mono Q column chromatography. The tubes were incubated for 10 minutes at 30°C, and subsequently 20 μl of bovine γ-globulin (20 mg/ml) and 200 μl of 24% polyethylene glycol (average 7,500

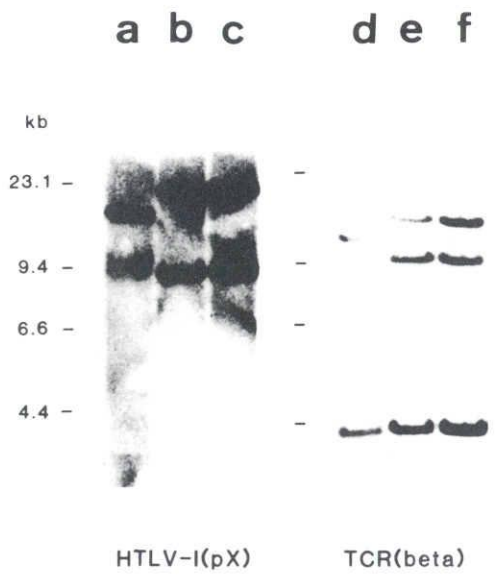



Fig. 1. Southern blot hybridization of DNA from ILT-Mat and TPA-Mat cells to gene probes of HTLV-I pX and T-cell receptor β chain. (**a**) TL-OmI, an HTLV-I-carrying T cell line as a positive control; (**b,e**) ILT-Mat cell line; (**c,f**) TPA-Mat cell line; (**d**) LCL-Kan, a human B cell lines as a negative control.

daltons) were added. After further incubation for 10 minutes at 0°C, the tubes were centrifuged for 10 minutes at 12,000 rpm. The supernatants were completely discarded, and the pellets were resuspended in 50 μl of PBS and placed on glass filters. The radioactivity was counted in a liquid scintillation counter. Nonspecific binding was measured in the presence of 10 μM unlabeled PDBu.

## RESULTS

Several T-cell lines established from peripheral blood leukocytes of adult T-cell leukemia patients carried the provirus genome of human T-cell leukemia virus type I (HTLV-I) and required IL-2 for growth as described previously (Sugamura et al., 1984). These cells proliferated transiently in response to TPA. One of these lines, ILT-Mat, was maintained in the growth medium containing IL-2 for more than a year. To examine whether or not IL-2 can be replaced by phorbol esters, we attempted to obtain a cell line dependent on TPA for growth from the culture of ILT-Mat cells ($1 \times 10^4$ cells/well) in medium containing 81 nM TPA in a 96-well microplate. After 3 weeks, proliferating cells were observed in about 80% of the wells, and a TPA-dependent cell line, named TPA-Mat, was established from a single well. Southern blot hybridization of cellular DNA with probes of T-cell receptor β (TCRβ) and HTLV-I genes demonstrated that the patterns of TCRβ and HTLV-I specific sequences observed in TPA-Mat and ILT-Mat were identical, confirming that TPA-Mat was derived from ILT-Mat (Fig. 1.) Cell-surface markers and expression of HTLV-I antigens were examined for TPA-Mat and its parental ILT-Mat cell line. Both cell lines were positive for CD2, Ia, IL-2 receptors

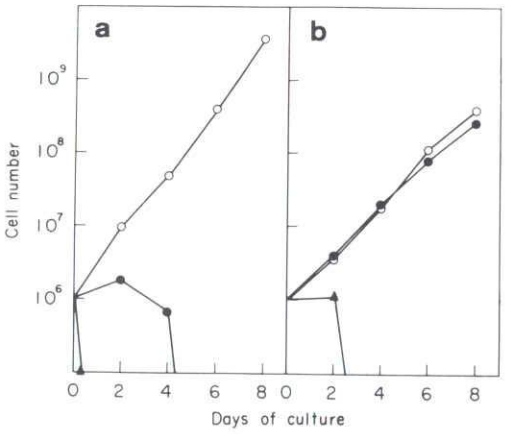



Fig. 2. Effects of IL-2 and phorbol esters on cell growth. ILT-Mat (**a**) and TPA-Mat (**b**) cells were cultured in the growth medium (▲), or growth medium containing 81 nM TPA (●) or 200 u/ml of IL-2 (○) in plastic dishes (Falcon no. 3002) and the medium was renewed every two days by centrifugation and readjusting the cell number to $2 \times 10^5$/ml or less. Viable cells were determined by the dye exclusion staining.

(p55 and p75), and 16% of ILT-Mat cells and 1% of TPA-Mat cells expressed HTLV-I antigens. They were negative for CD3, CD4, and CD8.

Figure 2 shows the growth curves of TPA-Mat and the parental ILT-Mat cells. Proliferation of TPA-Mat cells was observed in the presence of either TPA or IL-2, with a doubling time of about 24 hr. The TPA-dependency of TPA-Mat was not lost during cultivation for more than a year in the presence of TPA. In contrast, the parental ILT-Mat cells proliferated in response to IL-2 but died within 6 days in the presence of TPA. When TPA-Mat cells were maintained in the presence of IL-2 for more than half a year, they did not lose their TPA-dependency, which suggests that TPA-Mat is a stable mutant derived from ILT-Mat. As shown in Figure 3, TPA-Mat cells proliferated dose-dependently with TPA, PDBu, and IL-2, respectively, and half maximal incorporation of $^3$H-thymidine ($^3$H-TdR) was detected with 1.4 nM TPA, 47 nM PDBu, and 3.4 U/ml of IL-2. 4α-Phorbol-12,13-didecanoate (4α-PDD), which is a phorbol ester that does not cause tumor promotion or PKC activation, did not induce proliferation of TPA-Mat cells. Similarly, growth of TPA-Mat cells was promoted by another phorbol ester, 4β-PDD, and nonphorbol ester tumor promoters such as teleocidin, aplysiatoxin, and mezerein, all of which are known to activate PKC (Table 1). However, other phorbol esters, such as 4-O-methyl-TPA and 4β-phorbol, which do not cause PKC activation did not promote growth of TPA-Mat cells (Table 1). Also examined were the cumulative or synergistic effects of PDBu and IL-2 on TPA-Mat cell growth. As shown in Table 2, the cumulative effects rather than syngergistic effects were observed.

Next, TPA-Mat cells were examined for their ability to produce IL-2 in response to TPA. $^3$H-TdR uptake of TPA-Mat cells was assayed in the presence of antihuman recombinant IL-2 monoclonal antibody, which can neutralize IL-2 activity. As shown in Figure 4, the antibody did not affect TPA-induced incorporation of

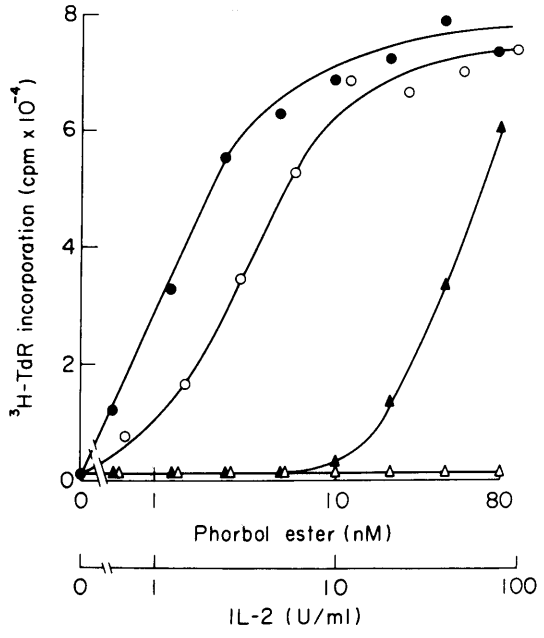



Fig. 3. Phorbol ester- or IL-2-dose dependent incorporation of $^3$H-TdR into TPA-Mat. TPA-Mat cells incubated for 48 hr in the growth medium containing the indicated doses of TPA (●), IL-2 (○), PDBu (▲) or 4α-PDD (△). During the last 4hr of incubation, $^3$HTdR was added, and $^3$H-TdR incorporated was measured.

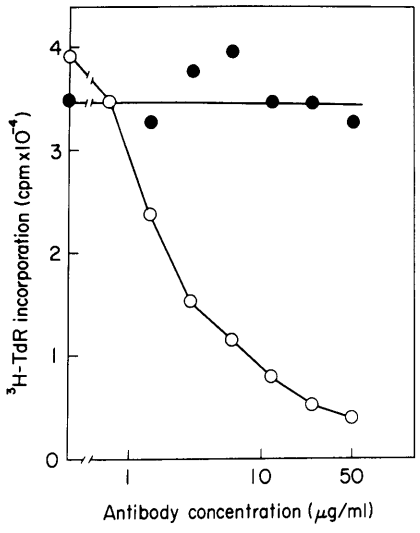



Fig. 4. Effects on anti-IL-2 monoclonal antibody on $^3$H-TdR incorporation into TPA-Mat. TPA-Mat cells were incubated for 48 hr in growth meduim containing serially diluted mouse monoclonal antibody to human recombinant IL-2 (L-61, obtained from Shionogi Pharm Co., Osaka, Japan) and 1.1 nM TPA ( ● ) or 1.5 u/ml of IL-2 ( ○ ) for 48 hr. During the last 4 hr of incubation, $^3$H-TdR was added, and $^3$H-TdR incorporation was measured.

TABLE 1. $^3$H-thymidine incorporation of TPA-Mat cells treated with phorbol esters and nonphorbol ester tumor promoters[1]

| Treated with | $^3$H-TdR incorporation (cpm) |
|---|---|
| TPA | 137,365 ± 4,267 |
| PDBu | 109,417 ± 3,533 |
| 4-O-methyl-TPA | 256 ± 58 |
| 4β-phorbol | 695 ± 385 |
| 4α-PDD | 374 ± 142 |
| 4β-PDD | 130,790 ± 6,817 |
| Mezerein | 132,275 ± 6,928 |
| Teleocidin | 138,848 ± 2,279 |
| Aplysiatoxin | 174,466 ± 2,433 |
| 10% FCS | 338 ± 305 |

[1]TPA-Mat cells precultured with the PDBu containing medium were washed and then assayed for $^3$H-TdR incorporation in the presence of 100 nM of the agents indicated.

TABLE 2. Cumulative effects between PDBu and IL-2 on $^3$H-thymidine incorporation of TPA-Mat cells[1]

| IL-2 (U/ml) | Incorporation of $^3$H-thymidine (cpm) with PDBu doses (in nM) | | | |
|---|---|---|---|---|
| | 0 | 1.5 | 3 | 6 |
| 0 | 6,363 | 24,500 | 47,237 | 94,718 |
| 2.5 | 47,742 | 75,364 | 108,120 | 132,128 |
| 5.0 | 87,645 | 120,095 | 149,921 | 162,093 |

[1]TPA-Mat cells were incubated for 48 hr in the growth medium containing the indicated doses of PDBu and IL-2. During the last 4 hr of incubaton, $^3$H-TdR was added, and $^3$H-TdR incorporated was measured.

$^3$H-TdR but inhibited its IL-2-induced incorporation. The supernatant of TPA-Mat cells cultured in the presence of TPA was also examined for IL-2 activity. Any significant activity of IL-2 in the supernatant was not detected by a $^3$H-TdR incorporation assay using a murine IL-2-dependent T-cell line (Table 3). Futhermore, no mRNA of IL-2 was detected in TPA-Mat or ILT-Mat cells in the presence of either TPA or IL-2 (Fig. 5) These results suggest that the TPA-dependent growth of TPA-Mat cells was not mediated by an autocrine mechanism involving IL-2.

We then investigated the PKC activities of the membrane and cytosol fractions of TPA-Mat and parental ILT-Mat cells (Table 4). The PKC activity in the cytosol fraction of ILT-Mat cells was detected and rapidly decreased after PDBu treatment, whereas the PKC activity in the membrane fraction increased inversely. In contrast, the PKC activity in TPA-Mat cells cultured with PDBu seemed to be down-regulated, as reported by others (Solanki et al., 1981; Jaken et al., 1981; Chida et al., 1986), because the activity was not detected in either the membrane or cytosol fraction at 0 hr of incubation with or without PDBu treatment. However, the activity became detectable in the cytosol fraction after preincubation without PDBu for at least 1 or 2 hr and was translocated from the cytosol fraction to the membrane fraction by PDBu treatment. As the PKC activity of TPA-Mat cells at 0 hr was undetectable, we tried to detect the activity with a more sensitive method. The number of TPA-Mat or ILT-Mat cells used for extraction was increased eightfold, and the extracts were fractionated with a Mono Q column (Fig. 6). A sharp peak of PKC activity in ILT-Mat cells was detected coincidentally with a main peak of $^3$H-PDBu binding activity. Similarly, a sharp peak of PKC and $^3$H-PDBu binding activities was detected in TPA-Mat cells, although the activities in TPA-Mat cells were less than one-seventh of those in ILT-Mat cells. We then examined the transcript of the α, β, and γ-genes of PKC in TPA-Mat cells by Northern blot

TABLE 3. IL-2 activity in culture supernatant of TPA-Mat cells[1]

| Samples | $^3$H-thymidine uptake (cpm) |
|---|---|
| Supernatant of TPA-Mat | 2,684 ± 349 |
| 14 nM TPA | 3,595 ± 392 |
| IL-2 (U/ml) | |
| 12 | 113,407 ± 3,062 |
| 6 | 63,298 ± 9,320 |
| 3 | 7,924 ± 1,431 |
| 1.5 | 2,646 ± 499 |
| 0 | 1,206 ± 132 |

[1]TPA-Mat cells were cultured for 3 days with the growth medium containing 81 nM TPA. The culture supernatant of TPA-Mat cells was concentrate sixfold by vacuum dialysis and assayed at the original concentration for IL-2 activity by the method of $^3$H-TdR incorporation by a murine IL-2-dependent cell line, CTLL-2. Serially diluted IL-2 and 14 nM TPA, which was estimated to remain in the supernatant of TPA-Mat cells, were used as controls.

hybridization with synthetic oligonucleotide probes complementary to the unique region of the gene. The results indicated that although PKC activity was down-regulated in the cells, the β-gene was significantly transcribed (Fig. 7).

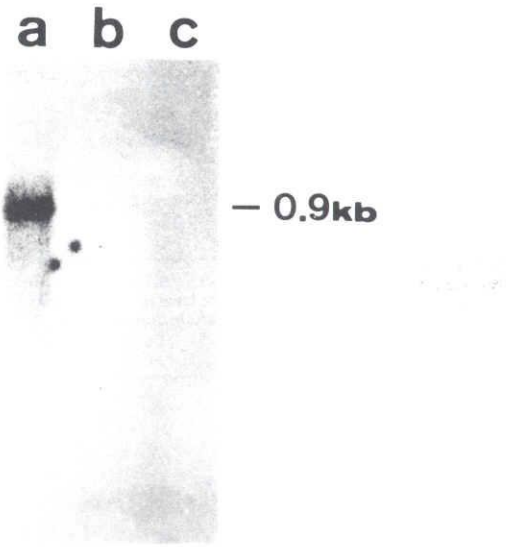



Fig. 5. Analysis of IL-2 mRNA in ILT-Mat and TPA-Mat. Normal peripheral blood leukocytes were cultured with growth medium containing 1% phytohemagglutinin M for 4 days (**a**). ILT-Mat cells were maintained in growth medium containing 200 u/ml of IL-2 (**b**). TPA-Mat cells were maintained in growth medium containing 81nM TPA (**c**). RNA samples isolated from these cells were hybridized with an IL-2 cDNA probe (Taniguchi et al., 1983).

## DISCUSSION

The present study shows that TPA or PDBu continuously induce the growth signal in TPA-Mat cells derived from an IL-2-dependent T-cell line, ILT-Mat. The induction of the growth signal by the phorbol esters was not mediated by an autocrine mechanism of IL-2 production, because TPA-Mat cells were found not to produce IL-2 in the presence of TPA. Furthermore, any other factor supporting their own growth was not detected in the supernatant of TPA-Mat cells cultured in the presence of TPA (data not shown). Therefore, the phorbol esters seem to induce the intracellular growth signal directly in TPA-Mat cells.

The growth signal was induced in TPA-Mat cells by phorbol esters, such as TPA and PDBu, that are known to be PKC activators, but not by phorbol esters, such as 4-0-methyl-TPA, 4β-phorbol, and 4α-PDD, which do not cause PKC activation. Growth of TPA-Mat cells was also promoted by nonphorbol ester tumor promoters such as teleocidin, aplysiatoxin, and mezerein, which also are known to activate PKC (Fujiki et al., 1984; Miyake et al., 1984). These observations suggest that PKC has an important role in the proliferation of TPA-Mat cells in the presence of the phorbol esters. A relationship between cell growth signaling and PKC activation has been proposed in various cell types. It was reported that TPA induced transient growth of IL-2-dependent murine T-cell lines (Kamber, 1986; Kim et al., 1986), and Farrar and Anderson (1985) and Farrar et al. (1985) also reported that IL-2 and IL-3 rapidly induced activation of PKC in an IL-2-dependent and IL-3-dependent murine cell lines, respectively, in accordance with all growth promotion. Moreover, other cell growth factors, such as PDGF and fibroblast growth factor (FGF), also are known to stimulate hydrolysis of phosphatidylinositol-4,5-bisphosphate to inositol-1,4,5-trisphosphate and diacylglycerol, known as an activator of PKC, in Swiss 3T3 fibroblasts (Berridge et al., 1984; Tsuda et al., 1985). Very recently, Pasti et al. (1986) obtained direct evidence that PKC activation by PDBu induces DNA synthesis in fibroblasts. They showed that microinjection of purified PKC into Swiss 3T3 fibroblasts pretreated with PDBu restored the mitogenic response of the cells to PDBu. These studies demonstrated the transient promotion of cell growth mediated by PKC in the presence of the phorbol ester. We have presented here direct evidence that the phorbol esters can induce continuous growth promotion of an IL-2-dependent T-cell line. TPA-Mat cells cultured with PDBu transcribed the β-gene of PKC, which is identical to the PKC II gene that is considered to be the gene for PKC found in lymphocytes (Knopf et al., 1986). Their PKC activity was down-regulated in the presence of phorbol esters, but activity was still detected when the cell extract was fractionated, and it increased with a short time preincubation in the absence of PDBu. Therefore, it is possible that the PKC was activated continuously by the phorbol esters and quickly degraded and also that this activity was sufficient for sustained induction of the growth signal in TPA-Mat cells. Intracellular calcium also has been suggested to participate in growth stimulation of T cells (Alford, 1970; Luckasen et al., 1974). Therefore, we preliminarily examined a possible involvement of calcium in phorbol ester-induced growth of TPA-Mat cells, but phorbol esters did not seem to increase intracellular calcium influx, and a calcium ionophore could not stimulate growth of TPA-Mat cells (data not shown).

The facts that TPA-Mat cells respond to IL-2 as well as to TPA and they cumulatively stimulated the growth of TPA-Mat cells, and also that IL-2 induced activation of PKC in the murine cell line (Farrar and Anderson, 1985), strongly suggest that PKC has a crucial role in signal transduction from IL-2 receptors triggered by IL-2. However, we have never detected IL-2-induced translocation of PKC from cytosol to

TABLE 4. Detection of PKC in cytosol and membrane fractions of ILT-Mat and TPA-Mat cells

| Cell line | Incubation time | PDBu treatment | Activity of PKC (pmol/min/mg protein) | |
|---|---|---|---|---|
| | | | Cytosol | Membrane |
| ILT-Mat | 0 hr | Untreated | 3,092 ± 71 | 285 ± 210 |
| | | Treated | 1,172 ± 6 | 2,942 ± 190 |
| TPA-Mat | 0 hr | Untreated | −110 ± 68 | −82 ± 69 |
| | | Treated | −86 ± 22 | −71 ± 28 |
| | 1 hr | Untreated | 41 ± 17 | −116 ± 78 |
| | | Treated | −19 ± 2 | 90 ± 44 |
| | 2 hr | Untreated | 243 ± 15 | −182 ± 163 |
| | | Treated | −57 ± 17 | 256 ± 16 |

ILT-Mat cells cultured with IL-2 and TPA-Mat cells cultured with PDBu were incubated in the growth medium without IL-2 or PDBu for the indicated times. Subsequently, they were untreated or treated with 200 nM PDBu for 10 min and then harvested for preparation of subcellular fractions. Subcellular fractionation was carried out as described in Materials and Methods.

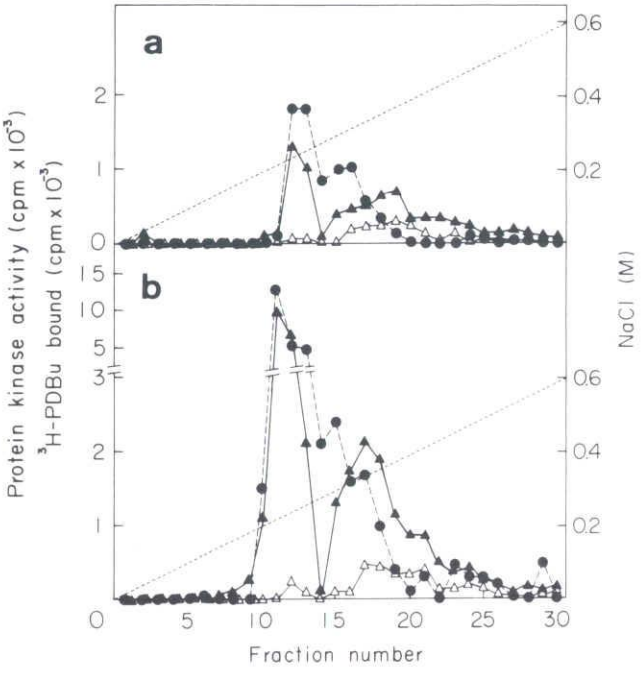



Fig. 6. Mono Q column chromatography of cell extracts. TPA-Mat (**a**) and ILT-Mat cell lysates (**b**) were applied to a Mono Q column and then eluted with a linear gradient (---) consisting of 0.0 – 0.6 M NaCl, and fractions of 1 ml were collected. 25μl aliquots of each fraction was used for the assay of PKC activity ( ● ), or $^{3}$H-PDBu binding in the presence ( ▲ ) or absence ( △ ) of phosphatidylserine.

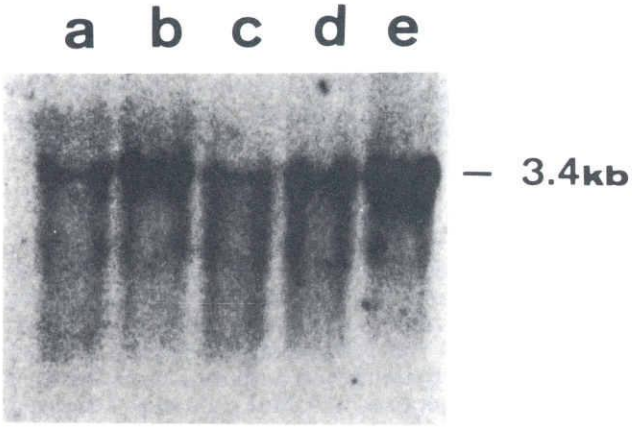



Fig. 7. Analysis of PKC β mRNA in ILT-Mat and TPA-Mat. ILT-Mat cells were cultured in growth medium containing 200 u/ml of IL-2 (**a**), and they were then incubated with 200 nM PDBu for 24 hr (**b**). TPA-Mat cells were cultured in growth medium containing 200 nM PDBu (**c**), and they were further incubated with 200 u/ml of IL-2 for two months (**d**). Normal peripheral blood leukocytes were cultured with the growth medium containing 1% phytohemagglutinin M and 200 u/ml of IL-2 for 7 days (**e**). RNA samples isolated from these cells were hybridized with a synthetic oligonucleotide probe (39 nucleotides) corresponding to nucleotide positions 548 and 586 of the anti-sense sequence of PKC β gene (Coussens et al., 1986).

membrane in TPA-Mat and ILT-Mat cells (data not shown), although the possibility cannot be ruled out that undetectable activation of PKC may contribute to the signal transduction. On the other hand, we recently obtained evidence that the phorbol ester-induced growth signaling could be different from the IL-2-induced growth signaling in TPA-Mat cells in respect to their sensitivity to adenosine 3′5′-monophosphate (cAMP) (Goto et al., manuscript in preparation). An increase of cAMP by treatment with forskolin completely inhibited the TPA-induced growth but slightly inhibited the IL-2-induced growth of TPA-Mat cells. These observations suggest a possibility that there are at least two different pathways for IL-2-induced signal transduction: one is a PKC-dependent pathway that is sensitive to cAMP, and the other is a PKC-independent pathway that is insensitive to cAMP. The comparative study between phorbol ester-dependent growth and IL-2-dependent growth may elucidate the fine mechanism of signal transduction for cell growth.

## ACKNOWLEDGMENTS

We thank Drs. H. Young (NCI, Frederick) and Y. Ito (Institute for Virus Research, Kyoto University , Japan) for helpful comments on this manuscript, and Drs. H. Fujiki (National Cancer Center Research Institute, Japan) and T. Taniguchi (Institute for Molecular and Cellular Biology, Osaka University, Japan) for generous gifts of tumor promoters and IL-2 cDNA probe, respectively.

This work was supported in part by grants-in-aid for Scientific Research and Cancer Research from the Ministry of Education, Science and Culture, and by grants from the Naito and Kudo Foundations.

## LITERATURE CITED


Alford, R.H. (1970) Metal cation requirements for phytohemagglutinin-induced transformation of human peripheral blood lymphocytes. J. Immunol., *104*:698–703.

Berenblum, I. (1941) The cocarcinogenic action of croton resin. Cancer Res., *1*:44–48.

Berridge, M.J., Heslop, J.P., Irvine, R.F., and Brown, K.D. (1984) Inositol triphosphate formation and calcium mobilization in Swiss 3T3 cells in response to platelet-derived growth factor. Biochem J., *222*:195–201.

Blumberg, P.M. (1980) In vitro studies on the mode of action of the phorbol esters, potent tumor promoters: Part 1. CRC Crit. Rev. Toxicol., *8*:153–197.

Blumberg, P.M., Driedger, P.E., and Rossow, P.W. (1976) Effect of a phorbol ester on a transformation-sensitive surface protein of chick fibroblasts. Nature, *264*:446–447.

Castagna, M., Takai, Y., Kaibuchi, K., Sano, K., Kikkawa, U., and Nishizuka, Y. (1982) Direct activation of calcium-activated, phospholipid-dependent protein kinase by tumor-promoting phorbol esters. J. Biol. Chem., *257*:7847–7851.

Chida, K., Kato, N., and Kuroki, T. (1986) Down regulation of phorbol diester receptors by proteolytic degradation of protein kinase C in a cultured cell line of fetal rat skin keratinocytes. J. Biol. Chem., *261*:13013–13018.

Coussens, L., Parker, P.J., Rhee, L., Yang-Feng, T.L., Chen, E., Waterfield, M.D., Francke, U., and Ullrich, A. (1986) Multiple, distinct forms of bovine and human protein kinase C suggest diversity in cellular signaling pathways. Science, *233*:859–866.

Depper, J.M., Leonard, W.J., Drogula, C., Kronke, M., Waldmann, T.A., and Greene, W.C. (1985) Interleukin 2 (IL-2) augments transcription of the IL-2 receptor gene. Proc. Natl. Acad. Sci. USA, *82*:4230–4234.

Farrar, W.L., and Anderson, W.B. (1985) Interleukin-2 stimulates association of protein kinase C with plasma membrane. Nature, *315*:233–235.

Farrar, W.L., Thomas, T.P., and Anderson, W.B. (1985) Altered cytosol/membrane enzyme redistribution on interleukin-3 activation of protein kinase C. Nature, *315*:235–237.

Fujiki, H., Tanaka, Y., Miyake, R., Kikkawa, U., Nishizuka, Y., and Sugimura, T. (1984) Activation of calcium-activated, phospholipid-dependent protein kinase (protein kinase C) by new classes of tumor promoters: teleocidin and debromoaplysiatoxin. Biochem. Biophys. Res. Commun., *120*:339–343.

Glisin, V., Crkvenjakov, R., and Byus, C. (1974) Ribonucleic acid isolated by cesium chloride centrifugation. Biochemistry, *13*:2633–2637.

Hecker, E. (1971) Isolation and characterization of the cocarcinogenic principles from croton oil. Methods Cancer Res., *6*:439–484.

Hedrick, S.M., Nielsen, E.A., Kavaler, J., Cohen, D.I., and Davis, M.M. (1984) Sequence relationships between putative T-cell receptor polypeptides and immunoglobulins. Nature, *308*:153–158.

Huberman, E., and Callaham, M.F. (1979) Induction of terminal differentiation in human promyelocytic leukemia cells by tumor-promoting agents. Proc. Natl. Acad. Sci. USA, *76*:1293–1297.

Inoue, J., Seiki, M., Taniguchi, T., Tsuru, S., and Yoshida, M. (1986) Induction of interleukin 2 receptor gene expression by $p40^{x}$ encoded by human T-cell leukemia virus type 1. EMBO J., *5*:2883–2888.

Jaken, S., Tashijian, A.H., Jr., and Blumberg, P.M. (1981) Characterization of phorbol ester receptors and their down-modulation in $GH_4C_1$ rat pituitary cells. Cancer Res., *41*:2175–2181.

Kamber, M. (1986) Phorbol myristate acetate-induced proliferation of an IL-2-dependent T-cell line: Action of PMA is independent of IL-2 and cannot be mimicked by diacylglycerols. Cell. Immunol., *102*:177–186.

Kaneshima, H., Hiai, H., Fujiki, H., Oguro, Y.B., Iijima, S., Sugimura, T., and Nishizuka, Y. (1983) Tumor promoter-dependent mouse leukemia cell line. Cancer Res., *43*:4676–4680.

Kim, D.-K., Otten, G., Moldwin, R.L., Dunn, D.E., Nau, G.J., and Fitch, F.W. (1986) PMA alone induces proliferation of some murine T cell clones but not others. J. Immunol., *137*:2755–2760.

Kitano, T., Go, M., Kikkawa, U., and Nishizuka, Y. (1986) Assay and purification of protein kinase C. Methods Enzymol., *124*:349–352.

Knopf, J.L., Lee, M.-H., Sultzman, L.A., Kriz, R.W., Loomis, C.R., Hewick, R.M., and Bell, R.M. (1986) Cloning and expression of multiple protein kinase C cDNAs. Cell, *46*:491–502.

Lehrach, H., Diamond, D., Wozney, J.M., and Boedtker, H. (1977) RNA molecular weight determinations by gel electrophoresis under denaturing conditions, a critical reexamination. Biochemistry, *16*:4743–4751.

Lotem, J., and Sachs, L. (1979) Regulation of normal differentiation in mouse and human myeloid leukemic cells by phorbol esters and the mechanism of tumor promotion. Proc. Natl. Acad. Sci. USA, *76*:5158–5162.

Luckasen, J.R., White, J.G., and Kersey, H. (1974 Mitogenic properties of a calcium ionophore, A23187. Proc. Natl. Acad. Sci. USA, *71*:5088–5090.

Maniatis, T., Fritsch, E.F., and Sambrook, J. (1982) Molecular Cloning. Cold Spring Harbor Laboratory, New York.

Miyake, R., Tanaka, Y., Tsuda, T., Kaibuchi, K., Kikkawa, U. and Nishizuka, Y. (1984) Activation of protein kinase C by non-phorbol tumor promoter, mezerein. Biochem. Biophys. Res. Commun., *121*:649–656.

Niedel, J.E., Kuhn, L.J., and Vandenbark, G.R. (1983) Phorbol diester receptor copurifies with protein kinase C. Proc. Natl. Acad. Sci. USA, *80*:36–40.

Nishizuka, Y. (1984) The role of protein kinase C in cell surface signal transduction and tumour promotion. Nature, *308*:693–698.

Ohtani, K., Nakamura, M., Saito, S., Noda, T., Ito, Y., Sugamura, K., and Hinuma, Y. (1987) Identification of two distinct elements in the long terminal repeat of HTLV-I responsible for maximum gene expression. EMBO J., *6*:389–395.

Pasti, G., Lacal, J.-C., Warren, B.S., Aaronson, S.A., and Blumberg, P.M. (1986) Loss of mouse fibroblast cell response to phorbol esters restored by microinjected protein kinase C. Nature, *324*:375–377.

Robb, R.J., Munck, A., and Smith, K.A. (1981) T cell growth factor receptors: Quantitations, specificity, and biological relevance. J. Exp. Med., *154*:1455–1474.

Robb, R.J., Rusk, C.M., Yodoi, J., and Greene, W.C. (1987) Interleukin 2 binding molecule distinct from the Tac protein: Analysis of its role in formation of high-affinity receptors. Proc. Natl. Acad. Sci. USA, *84*:2002–2006.

Rozengurt, E., Rodriguez-Pena, A., Coombs, M., and Sinnett-Smith, J. (1984) Diacylglycerol stimulates DNA synthesis and cell division in mouse 3T3 cells: Role of $Ca^{2+}$-sensitive phospholipid-dependent protein kinase. Proc. Natl. Acad. Sci. USA, *81*:5748–5752.

Sharon, M., Klausner, R.D., Cullen, B.R., Chizzonite, R., and Leonard, W.J. (1986) Novel interleukin-2 receptor subunit detected by cross-linking under high-affinity conditions. Science, *234*:859–863.

Sibley, D.R., Benovic, J.L., Caron, M.G., and Lefkowitz, R.J. (1987) Regulation of transmembrane signaling by receptor phosphorylation. Cell, *48*:913–922.

Smith, K.A., and Cantrell, D.A. (1985) Interleukin 2 regulates its own receptors. Proc. Natl. Acad. Sci. USA, *82*:864–868.

Solanki, V., Slaga, T.J., Callaham, M., and Huberman, E. (1981) Down regulation of specific binding of [20-$^{3}$H] phorbol 12,13-dibutylate and phorbol ester-induced differentiation of human promyelocytic leukemia cells. Proc. Natl. Acad. Sci. USA, *78*:1722–1725.

Southern, E.M. (1975) Detection of specific sequences among DNA fragments separated by gel electrophoresis. J. Mol. Biol., *98*:503–517.

Sugamura, K., Fujii, M., Kannagi, M., Sakitani, M., Takeuchi, M., and Hinuma, Y. (1984) Cell surface phenotypes and expression of viral antigens of various human cell lines carrying human T-cell leukemia virus. Int. J. Cancer, *34*:221–228.

Sugamura, K., Fujii, M., Kobayashi, N., Sakitani, M., Hatanaka, M., and Hinuma, Y. (1984) Retrovirus-induced expression of interleukin 2 receptors on cells of human B-cell lineage. Proc. Natl. Acad. Sci. USA, *81*:7441–7445.

Taniguchi, T., Matsui, H., Fujita, T., Takaoka, C., Kashima, N., Yoshimoto, R., and Hamuro, J. (1983) Structure and expression of a cloned cDNA for human interleukin-2. Nature, *302*:305–310.

Teshigawara, K., Wang, H.-M., Kato, K., and Smith, K.A. (1987) Interleukin 2 high-affinity receptor expression requires two distinct binding proteins. J. Exp. Med., *165*:223–238.

Touraine, J.-L., Hadden, J.W., Touraine, F., Hadden, E.M., Estensen, R., and Good, R.A. (1977) Phorbol myristate acetate: A mitogen selective for a T-lymphocyte subpopulation. J. Exp. Med., *145*:460–465.

Tsuda, T., Kaibuchi, K., Kawahara, Y., Fukazaki, H., and Takai, Y. (1985) Induction of protein kinase C activation and $Ca^{2+}$ mobilization by fibroblast growth factor in Swiss 3T3 cells. FEBS Lett., *191*:205–210.

Tsudo, T., Kozak, R.W., Goldman, C.K., and Waldmann, T.A. (1986) Demonstration of a non-Tac peptide that binds interleukin 2: A potential participant in a multichain interleukin 2 receptor complex. Proc. Natl. Acad. Sci. USA, *83*:9694–9698.

Weinstein, I.B., and Wigler, M. (1977) Cell culture studies provide new information on tumour promoters. Nature, *270*:659–660.

Weinstein, I.B., Wigler, M., and Pietropaolo, C. (1977) In: Origins of Human Cancer. H.H. Hiatt, D.J. Watson, and A.J. Winsten, eds. Cold Spring Harbor Laboratory, New York, pp. 751–772.

Yeung, Y.G., Jubinsky, P.T., Sengupta, A., Yeung, D.C.Y., and Stanley, E.R. (1987) Purification of the colony-stimulating factor 1 receptor and demonstration of its tyrosine kinase activity. Proc. Natl. Acad. Sci. USA, *84*:1268–1271.

VIROLOGY 169, 447–451 (1989)

# SHORT COMMUNICATIONS

## The Oncogenicity of Avian Adenoviruses

## IV. Confirmatory Evidence for Recombination between Viral and Cellular DNA Sequences and Repetition of the Recombinant in Cells of a Tumor Line

HIROSHI YASUE,*,[1] MASASHI IWAMI,*,[2] YOSHINAO KOIDE,*,[3] EIICHI OHTSUBO,† AND MASAHIDE ISHIBASHI*,[4]

*Laboratory of Viral Oncology, Research Institute, Aichi Cancer Center, Chikusa-ku Nagoya 464, Japan, and †Institute of Applied Microbiology, University of Tokyo, Bunkyo-ku, Tokyo 113, Japan

*Received August 30, 1988; accepted December 7, 1988*

On the basis of hybridization analyses with viral DNA fragments as probes, we previously concluded that, in DNAs of many fowl adenovirus type 1-induced rodent tumors, units consisting of viral DNA segments flanked on both sides by cellular DNA segments are repeated more than 100 times per diploid genome of cell. Molecular cloning of virus–cellular junction fragments and analyses of their sequence arrangement in the present study confirmed that the previous conclusion was basically correct. 

In many lines of tumors and transformed cells induced by fowl adenovirus type 1 (FAV-1) (*1*), a certain length of viral DNA segment is repeated more than 100 times per diploid genome of cell (*2, 3*). The length is rather uniform in a tumor (or transformed cell) line, though it varies from one tumor line to another. Southern blot hybridization analyses of DNAs of these lines using viral DNA segments as probes led us to conclude that the viral DNA segment at its termini undergoes recombination with cellular DNA sequences, and that the viral DNA segment together with flanking cellular DNA sequences as a unit is repeated up to 100 times or more. The study on a rat tumor line, RC13, for instance, predicted that there are two types of major virus–cellular repetition unit (Fig. 1A), and that the *Eco*RI cleavage would produce ca. 160 copies of a 10-kbp fragment as the major right virus–cellular junction fragment (MRJ), ca. 120 copies of a 10-kbp fragment and 40 copies of a 15-kbp fragment as two major left virus–cellular junction fragments (abbreviated to MLJ1, and MLJ2, respectively) per diploid genome. The present study aims to confirm the previous conclusion by attempting to examine whether one can molecularly clone the predicted MRJ and MLJs from RC13 DNA.

RC13 cell DNA was cleaved with *Eco*RI, and the resulting DNA fragments were cloned in Charon 4A phages as described previously (*4*). The resulting phages were screened by hybridization (*5*) using as probes viral DNA fragments, *Eco*RI-A and *Eco*RI-F labeled by nick-translation (*2*), since their portions were predicted to be a part of MRJ and MLJs, respectively. Thus, we picked five and six phage clones likely to contain the right and left junction fragments (JFs), respectively. The agarose gel electrophoresis of *Eco*RI-digested DNAs of the above five clones revealed that four of them had inserted fragments showing the same mobility as the predicted MRJ (data not shown). Inserts in the four clones were cleaved with *Bgl*II and with *Pst*I and analyzed by Southern blot hybridization (*2*) using the insert of one of the clones as a probe. All the inserts provided quite similar profiles of the autoradiogram (data not shown), indicating that all the inserts have the same sequence arrangement. Therefore, one of the four inserts was transferred to pBR325 for easier handling and designated as a candidate for MRJ (abbreviated to CMRJ). Similar analyses on six phage clones for the left JF indicated that inserts of five clones have the same sequence arrangement (data not shown). The insert of one of the five was designated as candidate for MLJ1 (abbreviated to CMLJ1).

The cleavage site of *Hin*dIII on CMRJ (Fig. 1B) was found to be the same as the predicted MRJ (Fig. 1A). Then, a segment of CMRJ delimited with *Eco*RI- and *Hin*dIII sites (marked by asterisks in Fig. 1B) was mapped by cleavage with each of enzymes, *Pst*I, *Mlu*I,

[1] Present address: Laboratory of Cell Genetics, National Institute of Animal Industry, Kukisaki-cho, Ibaraki-ken 305, Japan.

[2] Present address: Department of Biology, Faculty of Science, Nagoya University, Chikusa-ku, Nagoya 464-01, Japan.

[3] Present address: Amano Pharmacy Co., Ltd., Kasugai-shi, Aichi-ken, Japan.

[4] To whom all the correspondence should be addressed.



0042-6822/89 $3.00

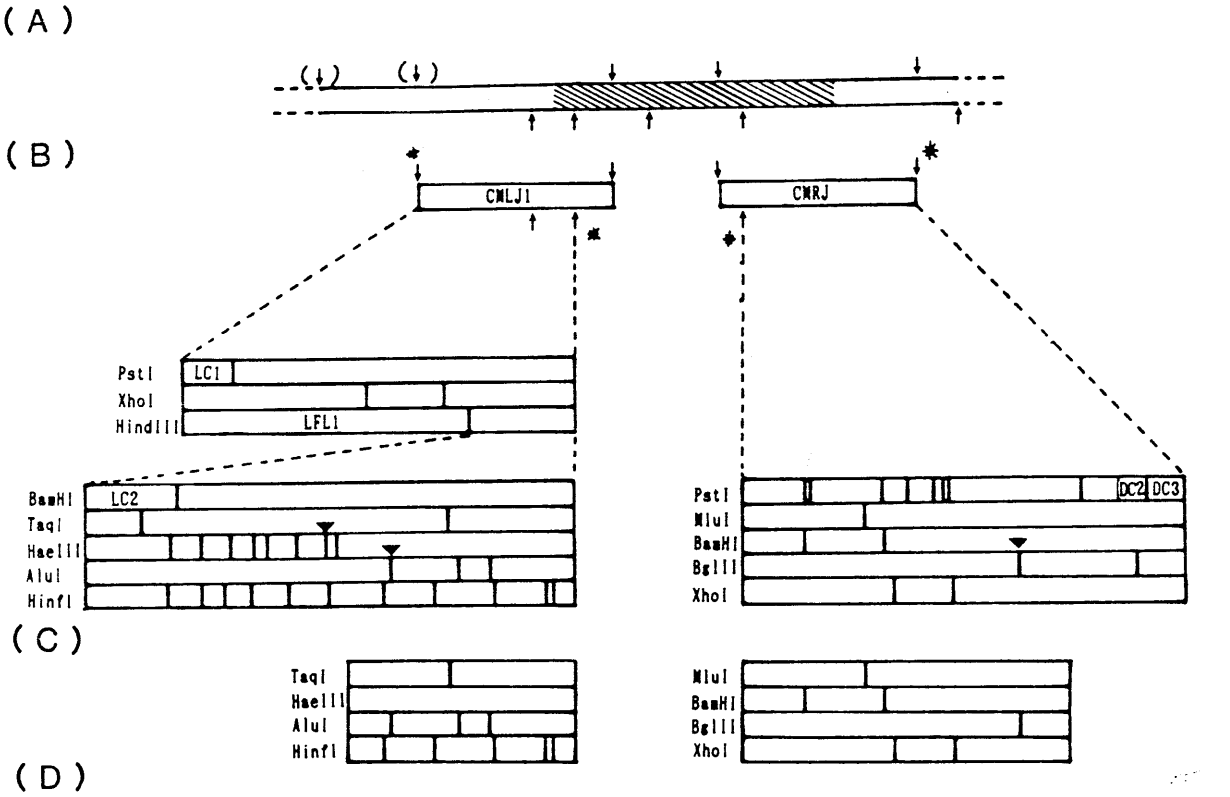



FIG. 1. The predicted major virus–cellular DNA repetition units, and cloned candidates for junction fragments. Descending arrows represent *Eco*RI sites; ascending arrows, *Hin*dIII sites. Asterisks above restriction sites in the figure are referred to in the text. (A) The cleavage map of two types of major virus–cellular DNA repetition units in RC13 cell DNA predicted previously (*3*): The hatched region represents viral sequence; the open region, cellular sequence. Two types of repetition units differ only in *Eco*RI sites shown in parentheses: The right one is for type I repetition unit and the left one for type II repetition unit. (B) The cleavage maps of cloned CMRJ and CMLJ1. Some segments delimited with *Hin*dIII and/or *Eco*RI were mapped more precisely with several restriction enzymes and presented in enlarged form. Procedures for mapping were described previously (*6*). (C) The cleavage maps of virion DNA fragments parts of which are predicted to be parts of JFs: The right side is *Hin*dIII-B fragment; the left side, *Hin*dIII-I fragment. (D) Base sequence in the vicinity of possible virus–cellular junction in CMRJ and CMLJ1, and the related DNA sequences. Actually, more than 270 consecutive bp extending rightward from the left *Bgl*II cleavage site (marked by arrowhead in (B)) of CMRJ and that of the corresponding virion DNA segment were determined. Similarly, all 306 consecutive bp of CMLJ1 ranging from an *Alu*I site to a *Hae*III site (marked by arrowheads in (B)), and base sequence of the corresponding virion DNA were analyzed. However, only a part of the base sequences of one of the two complementary strands each extending on both sides from the possible virus–cellular junction are presented. Bases are represented by the first letter of their spelling. Uppercase letters represent the bases of CMRJ and CMLJ1; lowercase letters, those of virion DNA, and of the variable region of the rat ribosomal DNA nontranscribed spacer (*11*). Underlining indicates AC-rich stretch

*Bam*HI, *Bgl*II, and *Xho*I (Fig. 1B). The comparison of cleavage maps for the latter four enzymes with those of *Hin*dIII-B segment of virion DNA (Fig. 1C) revealed that the segment extending from the left end to the *Bgl*II site marked by the arrowhead in Fig. 1B has the same cleavage pattern as the virion DNA segment, whereas the remaining segment does not. Therefore, the base sequences on the right side of the above *Bgl*II site and of the corresponding site of the virion DNA were determined and compared. The comparison revealed that both are identical in the first 146-bp sequence counted from the *Bgl*II site (asterisked in Fig. 1D), but that they have no significant homology in the more distal sequence. In this distal part of the CMRJ sequence, there is an AC-rich stretch (underlined in Fig. 1D) of 67 bases. A computer-aided search (*9*) of the rodent database in GenBank revealed that four genes such as murine-labile 4.5 S RNA contain sequences very similar to the above AC-rich sequences (initial score >170), and that seven genes such as rat elastase I gene contain sequences very similar to the TG-rich opposite strand. However, the sequence lying between the asterisked and the underlined did not show a similar high score for any gene in the database. These results may be interpreted as follows. The segment of CMRJ extending from its left end to the 146th base counted rightward from the above *Bgl*II cleavage site is derived from viral DNA, and the remaining segment is derived from the cellular DNA segment which includes an unidentified cellular gene containing an AC:TG-rich sequence.

Similarly, the *Hin*dIII cleavage sites on CMLJ1 (Fig. 1B) were found to be the same as those of predicted MLJ1 (Fig. 1A). Then, the DNA segment delimited with the *Eco*RI site and the *Hin*dIII site (marked by asterisks in Fig. 1B) was mapped by *Pst*I and *Xho*I cleavage. The segment delimited with two *Hin*dIII sites was more precisely mapped by cleavage with each of five enzymes (Fig. 1B). Comparison of maps for four enzymes with those of the *Hin*dIII-I segment of virion DNA (Fig. 1C) indicates that the *Hin*dIII-delimited segment of CMLJ1 consists of the part showing the same cleavage map as the *Hin*dIII-I fragment of virion DNA and the remaining part. Therefore, the base sequence was analyzed for the small subfragment of CMLJ1 extending from the *Alu*I site to the *Hae*III site (indicated with arrowheads in Fig. 1B), and for the corresponding segment of virion DNA. The comparison reveals that the contiguous 192 bp leftward from the *Alu*I site (asterisked in Fig. 1D) are identical to those of the virion DNA sequence. The computer-aided search of the rodent database found a quite high similarity (initial score 174; optimized score 304; marked by ::: in Fig. 1D) between the more leftward remaining sequence of CMLJ1 and the variable (or polymorphic) region of rat ribosomal DNA nontranscribed spacer (*11*), which is located ca. 300 bp upstream from the initiation site of transcription of 45 S rRNA precursor. The result can be interpreted as follows. The segment of CMLJ1 extending from its right end to the 192nd bp counted leftward from the *Alu*I cleavage site is the same as the leftmost *Eco*RI-cleaved virion DNA fragment missing only 4 bp of the left end, while the remaining is derived from the cellular segment which contains the above-mentioned variable region.

As a way to corroborate the above interpretation for CMRJ and CMLJ1, *Pst*I-produced subfragments of CMLJ1, named DC2 (0.6 kbp) and DC3 (0.7 kbp), and subfragments of CMLJ1, designated LFL1 (6 kbp) and LC2 (0.4 kbp) shown in Fig. 1B, were cloned, and their derivation was examined by their hybridizability with the dot-blotted rat embryo DNA and with similarly treated virion DNA. They were found to be hybridized with the rat embryo DNA, but not with the virion DNA

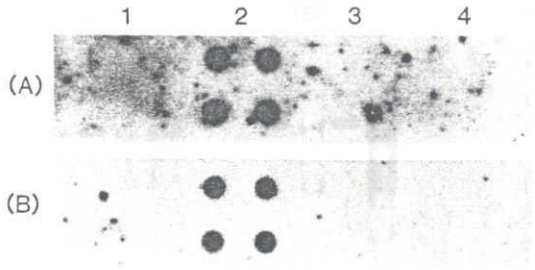



FIG. 2. Hybridizability of $^{32}$P-labeled subfragments of CMRJ or CMLJ1 with dot-blotted rat embryo DNA or virion DNA. Each DNA sample, after heat denaturation, was dot-blotted in quadruplicate on nitrocellulose filters, so that one can easily distinguish dots due to genuine hybridization response from those due to the background noise: (area 1) 1 μg of *Bacillus subtilis*; (area 2) 1 μg of rat embryo; (area 3) 0.3 μg of CELO virion; (area 4) 1 μg of *Micrococcus lysodeikticus*. Filters were baked at 80° *in vacuum* and were treated for hybridization (A) with a $^{32}$P-labeled equimolar mixture of DC2 and DC3, and (B) with a similarly prepared mixture of LFL1 and LC2 as probes. The hybridization conditions were the same as those described previously (*2*).

of 67 bases. (*) Stretch of base sequence found to be identical between CMRJ (or CMLJ1) and virion DNA. The possible junction in CMRJ corresponds to in between the 146th and 147th bp, counted rightward from the left *Bgl*II cleavage site, and that in CMLJ1, in between the 192nd and 193rd bp, counted leftward from the above *Alu*I site. (:::) Indicates the match between bases in CMLJ1 sequence and those of the variable region made by the computer-aided optimization (*9*), and (----) shows a gap fictionally produced by it. Base sequencing was done for both strands of each fragment by either one of the two methods described previously (*7, 8*). The data for the left-terminal two bases of virion DNA are according to Shinagawa *et al.* (*10*).

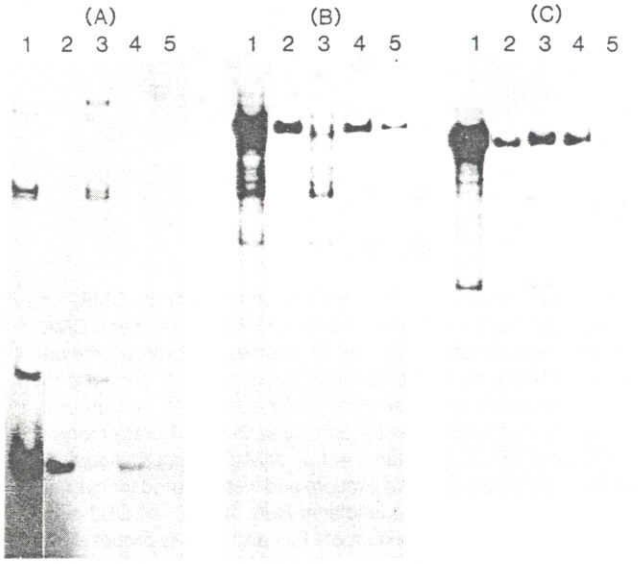



FIG. 3. Southern blot hybridization of *Pst*I- or *Eco*RI-cleaved RC13 DNA, rat embryo DNA, and CMRJ (or CMLJ1) with $^{32}$P-labeled DC2 (or LC1). (A) 5 μg and 0.5 μg of RC13 DNA (lanes 1 and 4, respectively), 5 μg of rat embryo DNA (lane 3), and 50 and 10 copy-equivalent CMRJ (lanes 2 and 5, respectively) were cleaved with *Pst*I, after *Bacillus subtilis* DNA as carrier was added to samples for lanes 2, 4, and 5, so that the final amount of DNA in each sample was 5 μg. Then, samples were electrophoresed on a 0.9% agarose gel at 40 V for 12 hr and blotted onto the nitrocellulose filter. The filter was treated for hybridization with $^{32}$P-labeled DC2, then for autoradiography. (B) Experimental conditions were the same as those in (A) except DNAs were cleaved with *Eco*RI. (C) Experimental conditions were the same as those in (B) except that 50 and 10 copy-equivalent of CMLJ1, in place of CMRJ, were subjected to the hybridization in lanes 2 and 5, respectively; also, LC1 was used as probe in place of RC2.

(Fig. 2). The result was confirmed by Southern hybridization tests using DC2, DC3, and LFL1 as probes, which showed these subsegments to be hybridized with several sizes of the *Pst*I- or *Eco*RI-cleaved rat embryo DNA fragments (see Figs. 3A and 3B for DC2; data not shown for DC3 and for LFL1), but with none of the *Hin*dIII-cleaved virion DNA fragments (data not shown), indicating that these subfragments are derived from moderately repeated sequences in rat DNA. All these results support the above interpretation.

A remaining requisite which can be experimentally confirmed is that number of repetition of CMRJ and CMLJ1 in RC13 DNA, when probed with the cellular sequences subcloned from themselves, should be the same as the number of repetitions of MRJ and MLJ1 predicted by our previous study with viral DNA segments as probes (*3*). Such experiments were done first with the above-described DC2 as a probe. Figure 3A shows results of the experiment in which examined DNAs were cleaved with *Pst*I. In the lane of 50 copy-equivalent of CMRJ as a control (lane 2), the band appears only at the site where DC2 should migrate, indicating that, even if any part of the DC2 sequence is repeated in the other part of CMRJ, its extent is very limited. In the lane of rat embryo DNA as another control (lane 3), a weak (10 copy-equivalent) band appears at the same site as the DC2 band in the above control, and stronger bands appear at several other sites closer to the electrophoretic origin. In the lane of RC13 DNA (lane 1), the mobility of most bands is the same as in the lane of rat embryo DNA, but the intensity of some bands is not the same as that of the corresponding bands in the lane of rat embryo DNA. Among others, the intensity of the band showing the same mobility as for DC2 band is remarkably stronger, and equivalent to ca. 200 copies per diploid genome. These findings indicate that DC2 is a constituent of normal rat cell DNA and is repeated ca. 10 and 200 times per diploid genome of rat embryo and of RC13 cells, respectively. Figure 3B shows results of a similar experiment in which examined DNAs were cleaved with *Eco*RI. In the lane of rat embryo DNA, among several other bands, a weak band (10 copy-equivalent or less) appears at the same site as the band in control CMRJ lane. In the lane of RC13 DNA, the mobility of most bands appears similar to that in the lane of rat embryo DNA, but the intensity of some bands appears stronger than that of the corresponding bands in the lane of rat embryo DNA. Among others, the band at the smae site as the control CMRJ band is as strong as ca. 200 copy-equivalent. It is worth remarking that the estimated number of copies (200) for the two above-mentioned strong bands shows rough agreement with the number (160) for MRJ estimated by our previous study using viral DNA as a probe, and the 200 copy-equivalent band in the lane of *Eco*RI-cleaved RC13 DNA appears at the same site as that of CMRJ. A similar experiment with DC3 as probe for DNAs cleaved with *Eco*RI gave results essentially similar to those shown in Fig. 3B (data not shown). All these results for numbers of repetitions strongly corroborate that CMRJ is the predicted MRJ.

Figure 3C shows the results of an analysis similar to Fig. 3B, except for having been probed with LC1, a subsegment of LFL1 (see Fig. 1B), which has been shown not to be hybridized with the other part of CMLJ1 (data not shown). In the lane of RC13 DNA, the strongest band (ca. 800 copy-equivalent) appears at the same site as the band in the control lane of CMLJ1. A strong band (100 copy equivalent) appears at the same site in the lane of rat embryo DNA. The difference between the copy number estimated for RC13 DNA and that for rat embryo DNA is as high as 700. However, other bands in the lane of RC13 DNA are also slightly more intense than those of the same mobility in the lane of rat embryo DNA. This fact seems attribut-

able to the probably inherent bias leading us to apply slightly more DNA in the former than in the latter. Even given this bias, the difference is no less than 300. It exceeds the increment of the copy number in RC13 DNA expected by the predicted number of repetition (ca. 120 copies per diploid genome) of MLJ1, and it thus indicates that other types of rearrangement take place in this tumor line, e.g., the cellular DNA sequence cleavable by *Eco*RI as a fragment which is the same size as CMLJ1 and which includes LC1 sequence is amplified very much. All these results for number of repetition are compatible with the notion that CMLJ1 is the predicted MLJ1.

As the selective hybridization marker for screening candidate phage clones incorporating MLJ1 or MLJ2, we used the viral DNA segment which was predicted to form a part of MLJ1 as well as MLJ2. However, no candidate for MLJ2 was found. At this moment, we think that the result reflects merely the statistical fluctuation of sampling. Thus, the existence of MLJ2, which we earlier concluded should be isolated, remains to be confirmed in future. Despite this incompleteness, the experimental results obtained here with CMRJ and CMLJ1 seem to confirm the previous conclusion in its major points.

Judging from the fact that no common sequence or sequence arrangement has been found in the vicinity of the recombination sites in several lines of human adenovirus-induced tumor cells so far extensively analyzed (*12, 13*), a more extensive analysis of the present JFs would not appear to provide supportive evidence for the hypothesis that there should be a sequence or sequence arrangement specific for the recombination. However, it is intriguing to consider that the recombination might take place preferentially on sites of host cell DNA which are located in or near transcriptionally active genes (*12*), in conjunction with a present conclusion that the left end of viral DNA may be linked to only ca. 300 bp upstream from the initiation site of transcription of 45 S rRNA precursor. To determine whether the present system supports the above hypothesis, we must do more extensive sequencing of CMLJ1 and molecular cloning of the corresponding preinsertion sequence from uninfected rat cells, as well as studies on their expression in cells.

## ACKNOWLEDGMENTS

The authors are grateful to Ms. S. Mori-Hori and Ms. T. Shibata-Sawaki for technical assistance. This work was in part supported by a grant-in-aid from the Princess Takamatsu Cancer Research Fund and a grant-in-aid for Cancer Research from the Ministry of Education, Science, and Culture, Japan.

## REFERENCES

1. ISHIBASHI, M., and YASUE, H., *In* "Adenoviruses" (H. S. Ginsberg, Ed.), pp. 497–562. Plenum, New York, 1984.
2. YASUE, H., ISHIBASHI, M., and FUJINAGA, K., *Virology* **106,** 361–373 (1980).
3. YASAU, H., and ISHIBASHI, M., *Virology* **16,** 99–115 (1982).
4. ENQUIST, L., and STERNBERG, N., *In* "Methods in Enzymology (R. Wu, Ed.), Vol. 68, pp. 281–298. Academic Press, New York, 1979.
5. BENTON, W. D., and DAVIS, R. W., *Science* **196,** 180–182 (1977).
6. SHIMADA, K., YOSHIDA, T., KUROISHI, T., and ISHIBASHI, M., *Biochim. Biophys. Acta* **740,** 169–178 (1983).
7. SÀNGER, F., NICKLEN, S., and COULSON, A. R., *Proc. Natl. Acad. Sci. USA* **74,** 5463–5467 (1977).
8. MAXAM, A. M., and GILBERT, W., *In* "Methods in Enzymology" (L. Grossman and K. Moldave, Eds.), Vol. 65 pp. 499–560 (author's comment). Academic Press, New York, 1980.
9. LIPMAN, D. J., and PEARSON, W. R., *Science* **227,** 1435–1441 (1985).
10. SHINAGAWA, M., ISHIYAMA, T., PADMANABHAN, R., FUJINAGA, K., KAMADA, M., and SATO, G., *Virology* **125,** 491–495 (1983).
11. YANG-YEN, F., and ROTHBLUM, L. I., *Nucleic Acid Res.* **14,** 5557 (1986).
12. LICHTENBERG, U., ZOCK, C., and DOERFLER, W., *J. Virol.* **61,** 2719–2726 (1987).
13. KUDO, S., SUGISAKI, H., TAKANAMI, M., YAMASHITA, T., ISHINO, M., SHIMIZU, Y., IIDA, Y., and FUGINAGA, W., *Virology* **148,** 133–145 (1986).

# RIBOSOMAL RNA GENES IN MYCOPLASMA CAPRICOLUM

A. Muto, F. Yamao, M. Sawada, M. Iwami, H. Hori and S. Osawa

Department of Biology, Faculty of Science, Nagoya University, Nagoya-shi, Aichi 464, Japan

Topics in Secondary Metabolism 1

*Bacillus subtilis*: Molecular Biology and Industrial Application

EDITED BY

BUNJI MARUO
*Nihon University, Tokyo, Japan*

HIROSHI YOSHIKAWA
*Osaka University, Osaka, Japan*

1989

KODANSHA LTD.
Tokyo

ELSEVIER
Amsterdam-Oxford-New York-Tokyo

## 2.2 Ribosomal RNA Genes in Mycoplasma Capricolum*

The mycoplasmas are small wall-less bacteria and parasitic in eukaryotes. Their genome size is about $5 \times 10^8$ daltons, which is about one-fifth that of *Bacillus subtilis*. We have estimated the number of genes in mycoplasma species is extremely low compared with other bacterial DNAs.[55]

We have recently isolated DNA clones including the rRNA genes from *Mycoplasma capricolum* ATCC27343 (Kid) and analyzed the organization and structure of the genes. The results suggest that the mycoplasma is phylogenetically related to the *Bacillus* group.

*Organization of rRNA genes:* Reflecting the small size of the genome, *M. capricolum* contains only two sets of genes for 16S, 23S and 5S rRNAs.[56] Since *B. subtilis* contains ten sets of rRNA genes,[57] the copy number of rRNA genes in the two species is roughly in proportion to their genome size.

The DNA segments carrying the rRNA gene clusters of *M. capricolum* were cloned in *Escherichia coli* plasmid pBR322, and analyzed by restriction endonuclease mapping, DNA-RNA hybridization and DNA sequencing.[58,59] Each of the two rRNA clusters designated as *rrnA* and *rrnB*, carries a set of genes for 16S, 23S and 5S rRNAs. The order and orientation of the rRNA genes in each cluster are (5′) 16S-23S-5S(3′) like in other prokaryotes, such as *E. coli*,[60] *B. subtilis*,[61] *Anacystis nidulans*[62] and *Rodopseudomonas capsulata*.[63] However, neither of the *M. capricolum* rRNA clusters includes any tRNA gene in the spacer between 16S and 23S rRNA genes. All seven rRNA operons of *E. coli*,[64–66] two rRNA gene clusters of *A. nidulans*[62] and two of ten rRNA gene clusters of *B. subtilis*[57] contain one or two spacer tRNA genes. Thus the overall organization of the *M. capricolum* rRNA gene clusters resembles that of the eight *B. subtilis* clusters that do not contain spacer tRNA genes.

The DNA sequences of a total 16S rRNA gene, the spacers and the 5′-

*Acknowledgement:

This work was supported by grants from the Ministry of Education, Science and Culture of Japan and from the Naito Science Foundation.

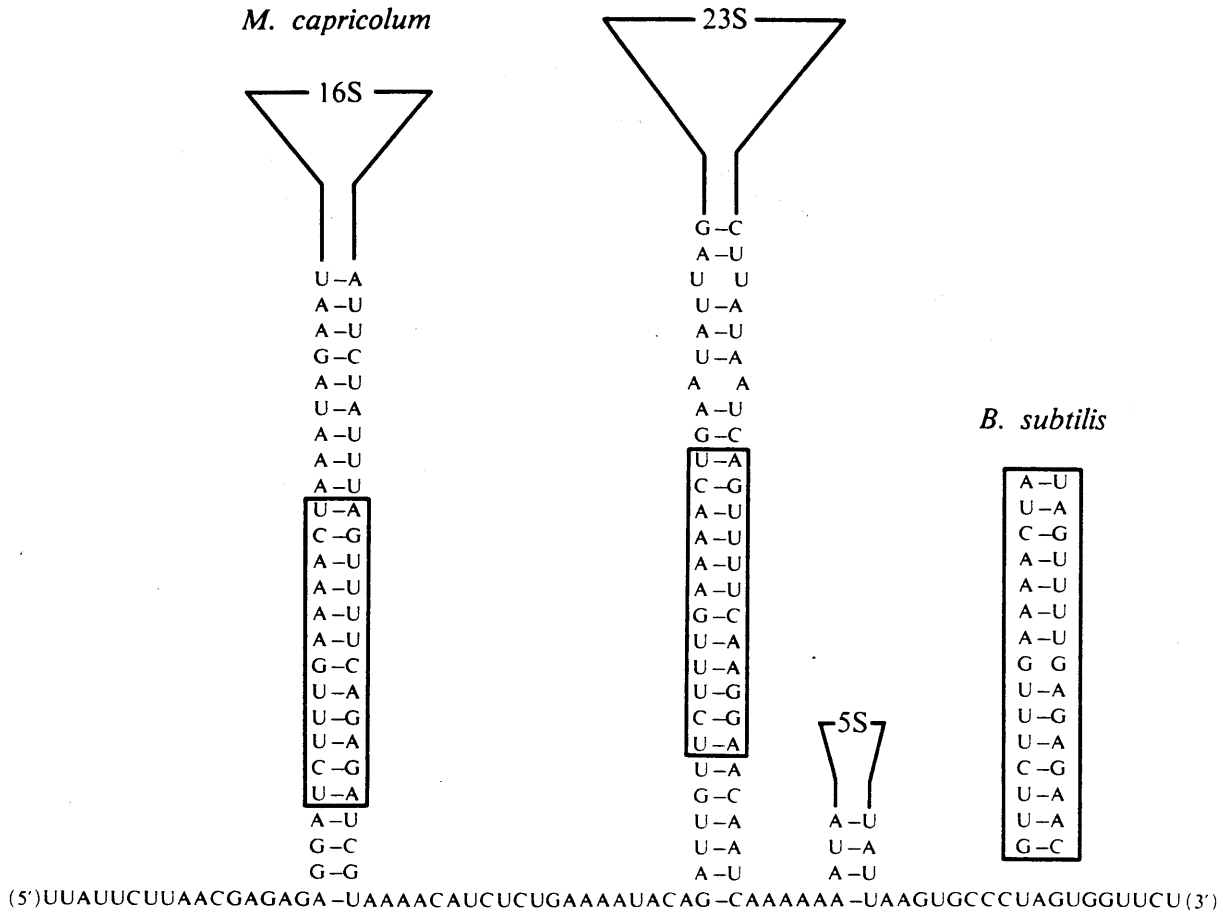



Fig. 2.3 Secondary structure of putative transcript of *rrnA*. The sequence is deduced from the DNA sequence of *M. capricolum rrnA*. The repeated sequences in the two stems are boxed. The repeated structure is also found in the stem of *B. subtilis rrnO*.[69]

and 3'-flanking regions of the cluster were determined. Comparison of the partial sequence of the 16S and 23S structure gene regions has revealed that the *M. capricolum* sequence is similar more to the *B. subtilis* (74%) than the *E. coli* sequences (69%).[57] The spacers and the flankings of the genes of *M. capricolum* are extremely poor in G and C (20%), as in the case of *B. subtilis*. Furthermore, a remarkable sequence homology is found in certain regions of spacer and flanking regions of the rRNA genes between *M. capricolum* and *B. subtilis*. Fig. 2.3 shows a possible secondary structure model of the putative transcript of *M. capricolum rrnA* with the flanking sequences. Two large stem structures can be folded between the 5'- and 3'-flanking sequences of the 16S and 23S rRNA sequences, respectively. Similar structures have been reported in the rRNA transcripts of *E. coli*[68] and *B. subtilis*.[69,70] It has been suggested that the stems may be possible substrates for processing enzymes during rRNA maturation. Moreover, the stretches of 12 base pairs are found to repeat in the 16S and 23S stems (sequences boxed in Fig. 2.3). Interestingly, the repeated stem structures with very similar sequences have also been observed in *B. subtilis*[69] (see

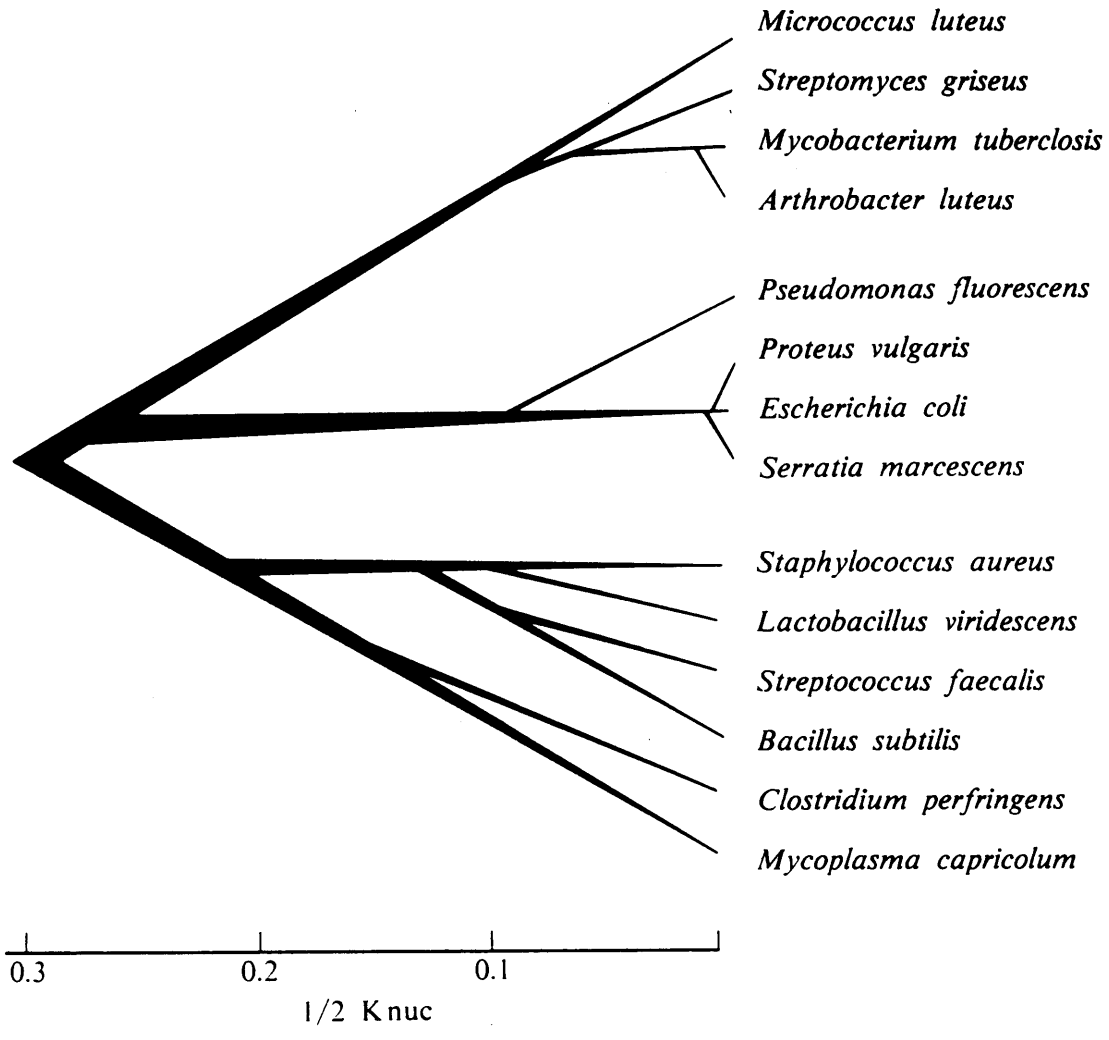



Fig. 2.4 Phylogenic tree of 5S rRNAs.
*Knuc* represents the rate of nucleotide substitution (see Ref. 72).

Fig. 2.3), but not in *E. coli.* The repeated structures in the two stems are located within the homologous regions in both RNAs. Ogasawara *et al.* have suggested that these regions may be the target sites for a processing enzyme.[69] This would mean that the processing signals of rRNA gene transcripts as well as the specificity of the processing enzyme are conserved between *M. capricolum* and *B. subtilis,* but not between these two species and *E. coli.* These observations suggest that *M. capricolum* is more closely related to *B. subtilis* than to *E. coli.*

*Nucleotide sequence of 5S rRNA:* The sequences of 5S rRNAs from about three hundred organisms have so far been determined and used for deducing their phylogenic relationships.[71,72] We have determined the total nucleotide sequence of *M. capricolum* 5S rRNA by sequencing both the purified RNA and the cloned DNA.[73] The RNA is 107 nucleotides long, the shortest of all the 5S rRNAs so far known. The sequence of *M. capricolum* 5S rRNA is more similar to that of the gram-positive bacteria than the gram-negative. Fig. 2.4. shows the phylogenic relationship of *M. capricolum* to other organisms deduced from the sequence homologies of

the 5S rRNAs (see Ref. 72). Walker *et al.* have sequenced 5S rRNA of *Spiroplasma* sp. BC3, indicating that spiroplasma is closely related to mycoplasma. Comparisons of mycoplasma tRNA sequences with other bacterial tRNAs also revealed the same relationship.[75]

(A. Muto, F. Yamao, M. Sawada, M. Iwami, H. Hori and S. Osawa)

## References


1. N.R. Smith, R.E. Gordon and F.E. Clark, *Aerobic Spore Forming Bacteria*, U.S. Department of Agriculture, Monograph no. 16 (1952).
2. R.E. Gordon, W.C. Haynes and C.H.-N. Pang, *The Genus Bacillus, Agricultural Handbook no. 427*, U.S. Department of Agriculture (1973).
3. T. Gibson and R.E. Gordon, *Bergey's Manual of Determinative Bacteriology* (ed. R.E. Buchanan and N.E. Gibbons), p.529, The Williams & Wilkins (1974).
4. N.E. Welker and L.L. Campbell, *J. Bacteriol.*, **94**, 1124 (1967).
5. N.E. Welker and L.L. Campbell, *J. Bacteriol.*, **94**, 1131 (1967).
6. J. Fukumoto, *J. Agric. Chem. Soc. Jpn.* (Japanese), **19**, 487 (1943).
7. T. Kaneda, *Bacteriol. Rev.*, **41**, 391 (1977).
8. J.N. Baptist, M. Mandel and R.L. Gherna, *Int. J. Syst. Bacteriol.*, **28**, 229 (1978).
9. T. Seki, T. Oshima and Y. Oshima, *Int. J. Syst. Bacteriol.*, **25**, 258 (1975).
10. D. Dubnau, I. Smith, P. Morell and J. Marmur, *Proc. Natl. Acad. Sci. USA*, **54**, 491 (1965).
11. J. Marmur, E. Seaman and J. Levine, *J. Bacteriol.*, **85**, 461 (1963).
12. J.R. Norris and D.H. Burges, *Entomophaga*, **10**, 41 (1965).
13. R.C.W. Berkeley and M. Goodfellow (eds.), *The Aerobic Endospore-Forming Bacteria, Classification and Identification*, Academic Press (1981).
14. J. Marmur and D. Lane, *Proc. Natl. Acad. Sci. USA*, **46**, 453 (1960).
15. P. Doty, J. Marmur, J. Eigner and C. Schildkraut, *Proc. Natl. Acad. Sci. USA*, **46**, 461 (1960).
16. C.L. Schildkraut, J. Marmur and P. Doty, *J. Mol. Biol.*, **3**, 595 (1961).
17. C.L. Schildkraut, K.L. Wierzchowski, J. Marmur, D.M. Green and P. Doty, *Virology*, **18**, 43 (1962).
18. H. Takahashi, H. Saito and Y. Ikeda, *J. Gen. Appl. Microbiol.*, **12**, 113 (1966).
19. T. Seki, C.-K. Chung, H. Mikami and Y. Oshima, *Int. J. Syst. Bacteriol.*, **28**, 182 (1978).
20. T. Seki, M. Minoda, J. Yagi and Y. Oshima, *Int. J. Syst. Bacteriol.*, **33**, 401 (1983).
21. J. Marmur, *J. Mol. Biol.*, **3**, 208 (1961).
22. J. Spizizen, *Proc. Natl. Acad. Sci. USA*, **44**, 1072 (1958).
23. P.W.J. Rigby, M. Dieckmann, C. Rhodes and P. Berg, *J. Mol. Biol.*, **113**, 237 (1977).
24. Y. Yoneda, K. Yamane, K. Yamaguchi, Y. Nagata and B. Maruo, *J. Bacteriol.*, **120**, 1141 (1974).
25. H. Kuhn, P.P. Fietzek and J.O. Lampen, *J. Bacteriol.*, **149**, 372 (1982).
26. S. Sawamura, *Bull. Coll. Agric. Tokyo Imp. Univ.*, **7**, 107 (1906).
27. A.G. O'Donnell, J.R. Norris, R.C.W. Berkeley, D. Claus, T. Kaneko, N.A. Logan and R. Nozaki, *Int. J. Syst. Bacteriol.*, **30**, 448 (1980).
28. P.S. Lovett and F.E. Young, *J. Bacteriol.*, **100**, 658 (1969).
29. F.J. Morgan and F.G. Priest, *J. Appl. Bacteriol.*, **50**, 107 (1981).
30. V.K. Krych, J.L. Johnson and A.A. Yousten, *Int. J. Syst. Bacteriol.*, **30**, 476 (1980).
31. L.K. Nakamura and J. Swezey, *Int. J. Syst. Bacteriol.*, **33**, 46 (1983).
32. L.K. Nakamura and J. Swezey, *Int. J. Syst. Bacteriol.*, **33**, 703 (1983).
33. T.Gibson, *Zentrabl. Bakteriol. Parasitenk. Infektionskr. Hyg. Abt. II*, **92**, 364 (1935).
34. R.E. Gordon, L.L. Hyde and J.A. Moore, Jr., *Int. J. Syst. Bacteriol.*, **27**, 256 (1977).
35. T. Kaneda, *J. Bacteriol.*, **95**, 2210 (1968).

36. H.J. Somerville and M.L. Jones, *J. Gen. Microbiol.*, **73**, 257 (1972).
37. T. Kaneko, R. Nozaki and K. Aizawa, *Microbiol. Immunol.*, **22**, 639 (1978).
38. K. Yoshimura, O. Yamamoto, T. Seki and Y. Oshima, *Appl. Environ. Microbiol.*, **45**, 1733 (1983) [Corrected version of the same paper; *Appl. Environ. Microbiol.*, **46**, 1268 (1983)].
39. J.W. Kronstad, H. Schnepf and H.R. Whiteley, *J. Bacteriol.*, **154**, 419 (1983).
40. G.A. Held, L.A. Bulla, Jr., E. Ferrari, J. Hoch, A.I. Aronson and S.A. Minnich, *Proc. Natl. Acad. Sci. USA*, **79**, 6065 (1982).
41. D. Dubnau, I. Smith and J. Marmur, *Proc. Natl. Acad. Sci. USA*, **54**, 724 (1965).
42. N. Harford and M. Mergeay, *Mol. Gen. Genet.*, **120**, 151 (1973).
43. T. Seki, H. Tsunekawa, K. Nakamura, K. Yoshimura and Y. Oshima, *J. Ferment. Technol.*, **57**, 488 (1979).
44. J.C. Copeland and J. Marmur, *Bacteriol. Rev.*, **32**, 302 (1968).
45. D.J. Henner and J.A. Hoch, *Microbiol. Rev.*, **44**, 57 (1980).
46. C.Goldthwaite, D.Dubnau and I. Smith, *Proc. Natl. Acad. Sci. USA*, **65**, 96 (1970).
47. N. Harford, J. Lepesant-Kejzlarovà, J.A. Lepesant, R. Hamers and R. Dedonder, *Microbiology-1976* (ed. D. Schlessinger), p.28, American Society for Microbiology (1976).
48. N. Harford and N. Sueoka, *J. Mol. Biol.*, **51**, 267 (1970).
49. J. Lepesant-Kejzlarovà, J.A. Lepesant, J. Walle, A. Billaut and R. Dedonder, *J. Bacteriol.*, **121**, 823 (1975).
50. M. Oishi, A. Oishi and N. Sueoka, *Proc. Natl. Acad. Sci. USA*, **55**, 1095 (1966).
51. D.J. Brenner, G.R. Fanning, K.E. Johnson, R.V. Citarella and S. Falkow, *J. Bacteriol.*, **98**, 637 (1969).
52. D.J. Brenner, G.R. Fanning, F.J. Skermann and S. Falkow, *J. Bacteriol.*, **109**, 953 (1972).
53. J.H. Crosa, D.J. Brenner, W.H. Ewing and S. Falkow, *J. Bacteriol.*, **115**, 307 (1973).
54. J. De Lay, I.W. Park, R. Tijtgat and J. Van Ermengem, *J. Gen. Microbiol.*, **42**, 43 (1966).
55. Y. Kawauchi, A. Muto, and S. Osawa, *Molec. Gen. Genet.*, **188**, 7 (1982).
56. M. Sawada, S. Osawa, K. Kobayashi, H. Hori and A. Muto, *Molec. Gen. Genet.*, **182**, 502 (1981).
57. H. Kobayashi and S. Osawa, *FEBS Lett.*, **141**, 161 (1982).
58. M. Sawada, A. Muto, M. Iwami, F. Yamao and S. Osawa, *Molec. Gen. Genet.*, **196**, 311 (1984).
59. M. Iwami, A. Muto, F. Yamao and S. Osawa, *Molec. Gen. Genet.*, **196**, 317 (1984).
60. C. Vola, B. Jarry and R. Rosset, *Molec. Gen. Genet.*, **153**, 337 (1977).
61. G.C. Stewart, F.E. Wilson and K. Bott, *Gene*, **19**, 153 (1982).
62. N. Tomioka, K. Shinozaki and M. Sugiura, *Molec. Gen. Genet.*, **184**, 359 (1981).
63. P.L. Yu, B. John, H. Falk and G. Drews, *Molec. Gen. Genet.*, **188**, 392 (1982).
64. E. Lund, J.E. Dahlberg, L. Lindahl, S.R. Jaskunas, P.P. Dennis and M. Nomura, *Cell*, **7**, 165 (1976).
65. T. Ikemura and M. Nomura, *Cell*, **11**, 779 (1977).
66. E.A. Morgen, T. Ikemura and M. Nomura, *Proc. Natl. Acad. Sci. USA*, **74**, 2710 (1977).
67. K. Loughney, E. Lund and J.E. Dahlberg, *Nucl. Acids Res.*, **10**, 1607 (1982).
68. R.A. Young and J.A. Steitz, *Proc. Natl. Acad. Sci. USA*, **75**, 3593 (1978).
69. N. Ogasawara, S. Moriya and H. Yoshikawa, *Nucl. Acids Res.*, **11**, 6301 (1983).
70. G.C. Stewart and K. Bott, *Nucl. Acids Res.*, **11**, 6289 (1983).
71. H. Hori and S. Osawa, *Proc. Natl. Acad. Sci. USA*, **76**, 381 (1982).
72. H. Hori, T. Itoh and S. Osawa, *Zbl. Bakt. Hyg. I. Abt. Orig.*, **C3**, 18 (1982).
73. H. Hori, M. Sawada, S. Osawa, K. Maruo and H. Ishikura, *Nacl. Acids Res.*, **9**, 5407 (1981).
74. R.T. Walker, E.T.J. Chelton, M.W. Kilpatrick, M.J. Rogers and J. Simmons, *Nucl. Acids Res.*, **10**, 6363 (1982).
75. M.W. Kilpatrick and R.T. Walker, *Nucl. Acids Res.*, **8**, 2783 (1980).

*(January, 1985)*

THE JOURNAL OF BIOLOGICAL CHEMISTRY
© 1989 by The American Society for Biochemistry and Molecular Biology, Inc.

Vol. 264, No. 13, Issue of May 5, pp. 7681-7685, 1989
Printed in U.S.A.

# cDNA Structure and Expression of Bombyxin, an Insulin-like Brain Secretory Peptide of the Silkmoth *Bombyx mori**

(Received for publication, January 5, 1989)

**Takashi Adachi‡§, Shigeharu Takiya‡, Yoshiaki Suzuki‡, Masafumi Iwami§, Atsushi Kawakami§, Susumu Y. Takahashi§¶, Hironori Ishizaki§‖, Hiromichi Nagasawa**, and Akinori Suzuki****

*From the ‡National Institute for Basic Biology, Myodaiji, Okazaki 444, Japan, the §Biological Institute, Faculty of Science, Nagoya University, Chikusa-ku, Nagoya 464, Japan, the ¶College of General Education, Yamaguchi University, Yamaguchi 753, Japan, and the **Department of Agricultural Chemistry, Faculty of Agriculture, the University of Tokyo, Tokyo 113, Japan*

**Bombyxin, previously referred to as 4K-prothoracicotropic hormone, is a brain peptide of the silkmoth *Bombyx mori*, the amino acid sequence of which shows considerable homology with vertebrate insulin family peptides. Two independent clones have been isolated from a *Bombyx* larval brain cDNA library by using a synthetic oligonucleotide probe, one with the complete coding region for preprobombyxin (λBb360) and the other covering the coding region, possibly for bombyxin, only partially (λBb204). λBb360 encodes preprobombyxin in the order of prepeptide/B-chain/proteolytic cleavage signal/C-peptide/proteolytic cleavage signal/A-chain. This domain organization of preprobombyxin is the same as that of preproinsulins, suggesting that the tertiary structure and posttranslational modification mechanism are conserved through the evolution of bombyxin and insulin. Genomic Southern hybridization analyses using this cDNA as probe suggest that the *Bombyx* genome contains multiple copies of bombyxin gene. Northern hybridization analyses indicate that the concentration of λBb360-type bombyxin mRNA in the bombyxin-producing cells is remarkably high ($2.8 \times 10^9$ molecules/μg of total RNA), without undergoing appreciable change during larval-pupal development.**

Prothoracicotropic hormone, an insect brain peptide, acts on the prothoracic glands to stimulate the synthesis and release of ecdysone necessary for growth, moulting, and metamorphosis (1, 2). The brain of the silkmoth *Bombyx mori* contains, in addition to its own prothoracicotropic hormone ($M_r$ ~22,000, Ref. 3), a 5-kDa peptide bombyxin that manifests the prothoracicotropic activity when tested with the saturniid moth *Samia cynthia ricini* (3, 4) and accordingly was referred to as 4K-prothoracicotropic hormone previously (5).

Bombyxin has been purified from *Bombyx* heads (4) and shown to consist of highly heterogeneous molecular species (6–10). Bombyxins are heterodimers consisting of A- and B-chains whose amino acid sequence shows considerable homology with vertebrate insulin family peptides (6, 7). The primary structure of bombyxin has been determined completely for three molecular species (7–9) and partially for three others (6, 10). These molecules differ from one another in a few replaced amino acid residues. The A- and B-chains are connected together through disulfide bonds in exactly the same way as in insulin (8). An immunohistochemical study using a monoclonal antibody raised against a synthetic fragment of bombyxin has revealed its localization in four pairs of mid-dorsal brain neurosecretory cells of *Bombyx* (5). Although the physiological function of bombyxin in *Bombyx* has remained unclarified, its insulin-like structure is thought to suggest some essential function in *Bombyx* growth or metabolism.

We have recently started studies using recombinant DNA technology for bombyxin cDNA and genomic DNA, with the aim of gaining insights into various molecular aspects of bombyxin including its biosynthesis through a precursor molecule, genetic background for the heterogeneity of molecular species, and transcriptional regulation during development. We describe here the cloning and characterization of bombyxin cDNAs from a *Bombyx* brain cDNA library. The structure of the deduced preprobombyxin confirms its domain organization conserved relative to preproinsulin, the fact of which was also expected from the simultaneously performed analysis of the genomic DNA fragment containing a similar sequence (11). Several hybridization analyses suggest that bombyxin multicopy genes supply a remarkably high concentration of bombyxin mRNA.

## EXPERIMENTAL PROCEDURES

*Materials*—*Bombyx mori* larvae were reared on an artificial diet (Takeda Pharmaceutical Co., Ltd., Tokyo, Japan or Nippon Nosanko Co., Yokohama, Japan) at 25 ± 1 °C under a photoperiod of 18 h light and 6 h dark. A total of 16,000 brains from the fourth- and fifth-instar larvae and pupae (J122 × C115) was used as the RNA source. RNA was also prepared from 16 pairs of the posterior silk glands of the fifth-instar larvae (Shun-Rei × Sho-Getsu). Three batches of a pair of the posterior silk glands of the fifth-instar larvae (J122 × C115) served as the source of genomic DNA. *Escherichia coli* Y1088 strain and λgt11 phage were provided by Dr. Yoshio Hamada (National Institute for Basic Biology, Okazaki, Japan). Various enzymes were purchased from Takara Shuzo Co., Ltd. (Kyoto, Japan), New England Biolabs Inc. (Beverly, MA), Boehringer Mannheim (Mannheim, Federal Republic of Germany), and Amersham Corp. Radioisotopes were purchased from Amersham Corp.

*RNA Extraction and Construction of cDNA Library*—Nucleic acids were extracted by phenol-SDS[1] from the brains after pulverization under liquid nitrogen in a mortar, and RNA was obtained by LiCl

* This work was partly supported by grants from the Ministry of Education, Science, and Culture of Japan, the Ishida Foundation, and the Mitsubishi Foundation. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

*The nucleotide sequence(s) reported in this paper has been submitted to the GenBank™/EMBL Data Bank with accession number(s) J04727 and J04728.*

‖ To whom correspondence should be addressed: Biological Institute, Faculty of Science, Nagoya University Chikusa-ku, Nagoya 464, Japan. Tel.: 52-781-5111, (ext. 2472).

[1] The abbreviation used is: SDS, sodium dodecyl sulfate.

precipitation. Oligo(dT)-primed cDNA was synthesized by cDNA synthesis system (Amersham Corp.) using poly(A)$^+$ RNA from the brains of the fourth- to fifth-instar larvae as template. cDNA was inserted into the *Eco*RI site of λgt11 DNA as described (12). Screening was carried out after amplification of the library.

*cDNA Screening and Analysis*—Employing the codon usage of *Bombyx* major plasma protein (13), a synthetic oligonucleotide 51-mer of the antisense DNA for bombyxin-II A-chain (the residues 4–20, Ref. 7) was prepared as a probe by Drs. Tadayuki Takeda and Masayuki Yamamoto (The University of Tokyo) as shown below.

```
Amino acid   C-Cys Tyr Ser Leu Leu Val Asp Val Ser Cys Pro Arg Leu Cys Cys Glu Asp-N
Antisense    5'-GCA GTA GCT GAG GAG GAC GTC GAC GCT GCA GGG TCT GAG GCA GCA CTC GTC-3'
```

Northern hybridization and plaque hybridization were carried out with $^{32}$P-end-labeled probe. Brain total RNA extracted by the above-mentioned method served for calculating the bombyxin mRNA concentration. Nucleotide sequences were determined by the M13 dideoxy method (14). For computer search of translation, hydropathy, and homology, GENETYX (Software Development Co., Tokyo, Japan) and DNASIS (Hitachi Software Engineering, Tokyo, Japan) programs were used.

*In Vitro Synthesis of Bombyxin RNA with T7 Polymerase and Northern Blotting*—After recloning of the cDNA insert in pTZ18R, bombyxin RNA was synthesized with T7 polymerase as described by Kumar *et al.* (15). After the addition of an appropriate amount of posterior silk gland total RNA as a carrier to the synthetic bombyxin RNA, RNA samples were glyoxalated and electrophoresed on a 1.2 or 1.4% agarose gel and then transferred to nylon membrane (biodyne A, PALL Ultrafine Filtration Co., New York) with 20 × SSC (3 M NaCl, 300 mM sodium citrate) as described (16).

*Genomic DNA Extraction and Southern Blotting*—Every sample of *Bombyx* genomic DNA was extracted from one pair of the fifth-instar posterior silk gland as described (17) except that 1.1 mg/ml proteinase K was employed during homogenate incubation with SDS. Genomic DNAs digested with various restriction enzymes (5.8 μg) were electrophoresed on 0.7% agarose gel and analyzed by Southern blot hybridization (18).

*Hybridization with Radiolabeled Probe and Washing*—Hybridization with oligonucleotide probe was carried out for more than 12 h at 22–25 °C in 50% formamide/4 × SSC (600 mM NaCl, 60 mM sodium citrate)/1 × Denhardt's solution (0.02% polyvinylpyrolidone, 0.02% Ficoll, 0.02% bovine serum albumin)/50 mM phosphate buffer (pH 5.95)/250 μg/ml sonicated salmon testis DNA. Filters were washed for more than 3 h at 60 °C in 2 × SSC/0.1% SDS ($T_m$, −30 °C). In the case of cDNA probe, hybridization solution contained 50% formamide/4–5 × SSC/1–5 × Denhardt's solution/50 mM phosphate buffer/250 μg/ml sonicated salmon testis DNA, and reactions were done overnight at 42 °C. Filters were washed for more than 3 h at 65 °C in 1 × SSC/0.1% SDS ($T_m$, −25 °C).

## RESULTS

*Isolation of Bombyxin cDNA from a Bombyx Larval Brain cDNA Library*—The condition of screening was first investigated by using the oligonucleotide as a probe for Northern hybridization of brain poly(A)$^+$ RNA. As shown in Fig. 1*A*, a hybridization band of about 0.6 kilobase pair was detected in the poly(A)$^+$ RNA from brain but not from posterior silk gland, suggesting the specific hybridization of bombyxin mRNA with the oligonucleotide probe. About $6 \times 10^4$ recombinant phages were then screened under the same condition, and 12 positives were revealed. Two of them, λBb360 and λBb204, were subjected to further analysis. In Fig. 1*B*, λBb360 cDNA was used as a probe in the Northern hybridization of the same filter as used in the experiment for Fig. 1*A*. The similar pattern of both hybridizations indicates that λBb360 cDNA is a complement of the mRNA that was specifically detected with the oligonucleotide probe.

*Structure of Bombyxin cDNAs and Predicted Proteins*—Fig. 2*A* shows the nucleotide sequences of λBb360 and λBb204 cDNAs and their predicted amino acid sequences. λBb360 cDNA obviously contains the entire coding region, because it has an open reading frame from +42 to +308 preceded by a stop codon at +30 to +32, four frames upstream from the initiator. By contrast, λBb204 cDNA appears to cover only a 3′ portion of the coding region.

The primary product deduced from λBb360 cDNA contains segments that resemble the A- and B-chains of naturally occurring bombyxins (*solid line-boxed* in Fig. 2*A*), with only minor amino acid residue replacement. The deduced sequences most closely resemble bombyxin-II, with precise identity for A-chain and 89% homology for B-chain.

The Gln at the N terminus of the presumed B-chain differs from pyroglutamate at the corresponding site of the naturally occurring bombyxins but is most probably converted to pyroglutamate posttranslationally (19). Between the A- and B-chains is found an intervening segment which is flanked by paired basic amino acid residues at the N terminus and a basic residue at C terminus (*broken line-boxed* in Fig. 2*A*), suggesting that the posttranslational proteolytic cleavage occurs at these sites; a single Arg is well documented as a proteolytic signal (20). Neighboring the N terminus of the B-chain a hydrophobic stretch is located. These intervening and hydrophobic sequences appear, from their structural features and organization, to be equivalent to the C-peptide and prepeptide of preproinsulin, respectively (21). To assign an appropriate length and hydrophobicity to a putative prepeptide, we consider Met-1 as the most likely initiator among three potential ones, Met-1, -10, and -15. Thus, a preprobombyxin deduced from λBb360 cDNA is an 89-residue protein with a $M_r$ of 9,978 having an overall structure of prepeptide/B-chain/proteolytic cleavage signal/C-peptide/proteolytic cleavage signal/A-chain which is exactly the same as that of preproinsulin (Fig. 3*B*). The hydropathy profiles (22) of preprobombyxin and preproinsulin are also similar to each other (Fig. 3*B*, *top* and *bottom*).

λBb204 cDNA encodes the entire A-chain (*solid line-boxed* in Fig. 2*A*) and, most probably, the C-peptide which is flanked at its C terminus by paired basic amino acid residues, and we presume that this cDNA was derived from a bombyxin mRNA. This cDNA differs from λBb360 cDNA in many aspects (Fig. 2*B*), indicating the existence of heterogeneous bombyxin mRNAs. The A-chain differs from λBb360 A-chain

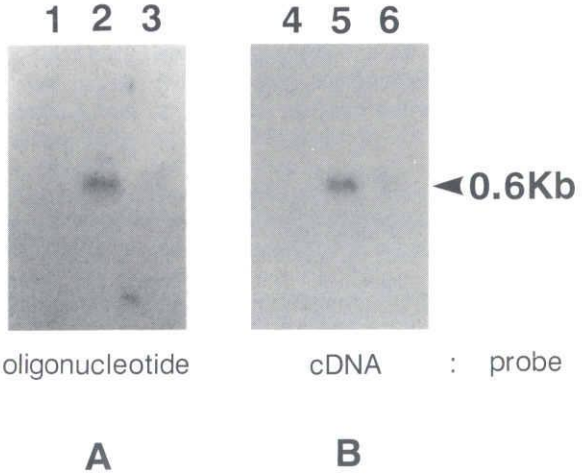
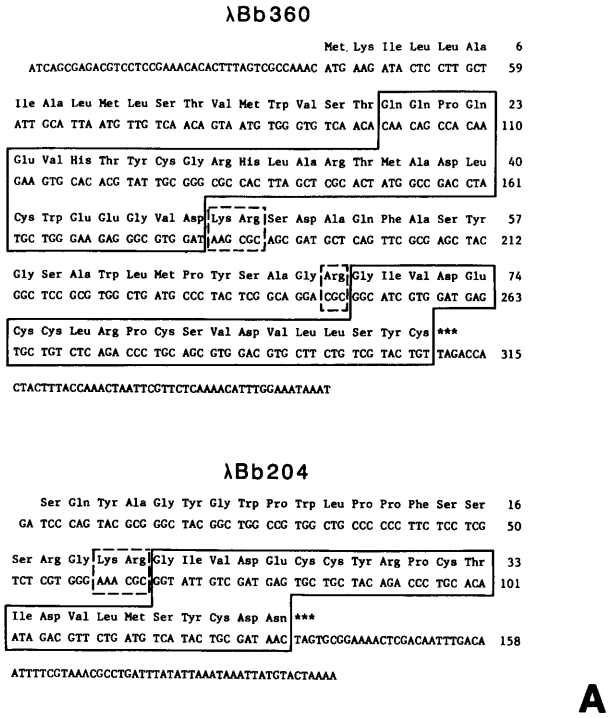



FIG. 1. **Northern hybridization analysis of bombyxin mRNA.** Posterior silk gland poly(A)$^+$ RNA (3.3 μg) (*lanes 1* and *4*), brain poly(A)$^+$ RNA (5.0 μg) (*lanes 2* and *5*), and brain total RNA (20 μg) (*lanes 3* and *6*) were electrophoresed and blotted. *A*, hybridization with $^{32}$P-end-labeled synthetic oligonucleotide. *B*, hybridization with $^{32}$P-multiprime-labeled λBb360 cDNA. *kb*, kilobase pair.



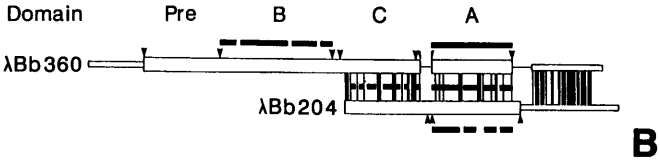




FIG. 2. *A*, nucleotide sequences and encoded amino acid sequences of cDNA inserts of λBb360 (*upper*) and λBb 204 (*lower*). Two domains forming mature bombyxin (A- and B-chain) and the putative proteolytic cleavage sites (Lys-Arg, Arg) are *boxed* with *solid* and *broken lines*, respectively. *B*, diagrammatic presentation of both cDNA structures. *Thick* and *thin open boxes* are coding and noncoding regions, respectively. Two *thin black lines* are introduced in λBb360; the first one, between C-peptide and A-chain, indicates a short deletion as compared with λBb204, and the second one is placed in order to align the polyadenylation signal with that of λBb204. *Vertical lines* between the two cDNAs connect the nonidentical nucleotides. *Horizontal black bars* at the *middle* represent identical amino acid residues between the two predicted peptides. *Horizontal black bars* at the *top* and *bottom* represent identical amino acid residues when compared with bombyxin-II (7). *Arrowheads* indicate borders between domains.

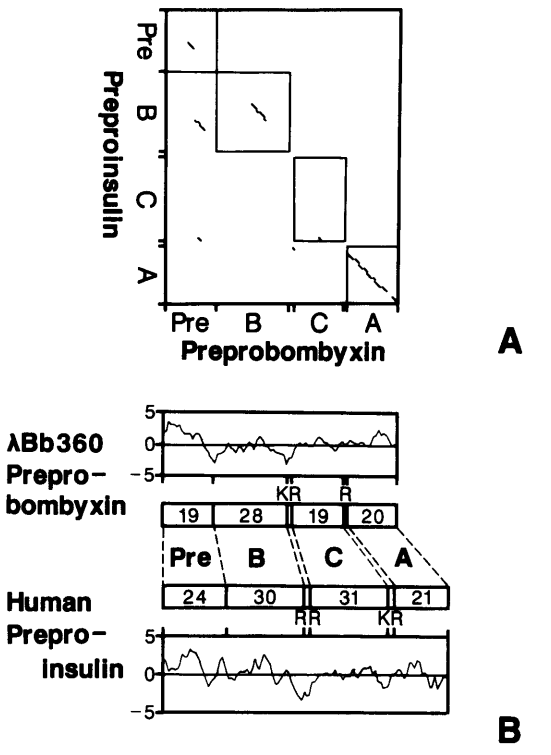



FIG. 3. **Comparison of structures between λBb360 preprobombyxin and human preproinsulin.** *A*, homology plot of amino acid sequences between λBb360 preprobombyxin and human preproinsulin produced by the program DNASIS (Hitachi Software Engineering, Tokyo). Span-length, 10; identities, 4. *B* (*middle*), schematic representation of domain organization of preprobombyxin encoded by λBb360 cDNA and human preproinsulin. *Boxes* and *numerals* represent the domains and the numbers of amino acid residues in each domain, respectively. Basic amino acids at the putative proteolytic cleavage sites are indicated by the one-letter code (*K* or *R*). *Top* and *bottom*, hydropathy profiles of both preprohormones using the method of Kyte and Doolittle (22). The values for every five sequential amino acids are calculated.

by 4 residues, and its C terminus is extended by 2 residues giving rise to a 22-residue A-chain contrasting to the 20-residue A-chain of all the bombyxins so far sequenced.

When nucleotide and/or amino acid sequences of the cDNAs characterized here were compared with the preprobombyxin gene isolated recently (11) and native bombyxins (6–10), it is evident that the homology is highest in the A-chain and only slightly lower in the B-chain; homology for the C-peptide is far less prominent; no significant homology is seen in the 3′-untranslated region. The same tendency in the degree of sequence homology among domains is also apparent when compared with insulin (Fig. 3*A*) and has been recognized among various insulins (23).

All of the 6 amino acid residues that are substituted compared to bombyxin-II (positions 8, 12, 13, and 17 of λBb204 A-chain and 5 and 25 of λBb360 B-chain) occupy the sites known to be variant from peptide analysis of native bombyxins (6–10). Another substitution of Met-18 for Leu-18 in λBb360 B-chain is a conservative substitution to retain the hydrophobic core inside the globular structure of bombyxins (10). Therefore, it may well be that λBb360 cDNA and possibly λBb204 cDNA also code for biologically active novel bombyxins.

*Genomic Southern Hybridization: Estimation of Bombyxin Gene Copy Number*—In view of the detection of two distinct cDNAs encoding bombyxin as described above and the presence of multiple molecular species of bombyxins as revealed by amino acid sequencing of the purified native peptides (6–10), it was anticipated that multiple bombyxin gene copies would exist. To test this possibility, the genomic DNA of *Bombyx* was analyzed by Southern hybridization using the λBb360 cDNA as probe. As shown in Fig. 4, multiple discrete hybridization bands were detected in every digest of genomic DNA. Taking the band intensity of five-copy λBb360 cDNA as 5.0, the total band intensities for respective lanes were estimated to range from 11 to 26 (Fig. 4, *lower*). As the genomic DNA was prepared from each individual separately, the possible restriction fragment length polymorphisms were minimized. Thus, it seems highly probable that there exist bombyxin gene copies in about these numbers, although the number might be less when the possible inclusion of pseudogenes is considered. It is also possible that the actual number of gene copies could be more since the probe presently used failed to hybridize with λBb204 cDNA under the same conditions (data not shown); only the genomic fragments containing the sequence identical or closely related to λBb360

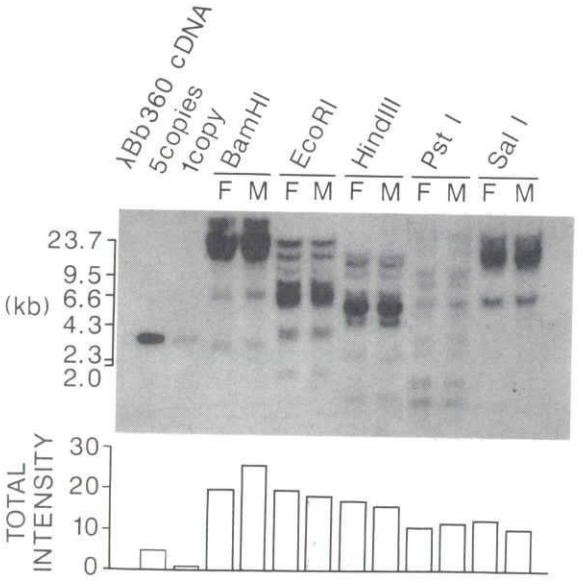



FIG. 4. **Southern hybridization analysis of genomic DNA of *Bombyx* (J112 × C115) probed with λBb360 cDNA.** *Upper,* in *left* two *tracks,* five and one copy equivalents of pBb360 (pTZ18R with λBb360 cDNA insert) linearlized by *Hin*dIII cut were loaded as the standard for copy number. Enzymes used for genomic DNA digestion are indicated above the autoradiogram. Female (*F*) and male (*M*) genomic DNAs were extracted from a single individual, respectively, and were dispensed in 8-μg aliquots for each restriction digestion except for the case of *Sal*I-male where a 5- and 3-μg portion were combined from two individuals. Out of these digests, 5.8 μg each were applied for the electrophoresis. Genome size of *Bombyx* is 0.52 pg/haploid. Hybridization probe was multiprimed $^{32}$P-labeled λBb360 cDNA. *Lower,* total densitometric band intensities for respective lanes taking the intensity for five-copy λBb360 cDNA as 5.0. Their relative indices are: five copy, 5.0; one copy, 0.8; *Bam*HI, 19.8 (*F*), 25.7 (*M*); *Eco*RI, 19.6 (*F*), 18.2 (*M*); *Hin*dIII, 17.1 (*F*), 16.0 (*M*), *Pst*I, 11.0 (*F*), 12.0 (*M*), *Sal*I, 12.8 (*F*), 10.5 (*M*), *kb,* kilobase pair.

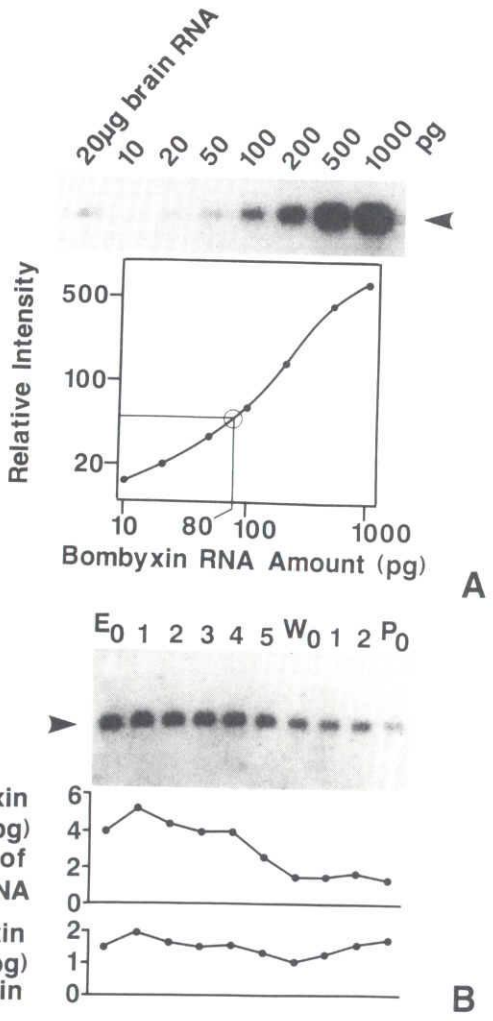



FIG. 5. **Quantification of bombyxin mRNA in *Bombyx* brain by the use of λBb360 cDNA.** *A,* quantification method. *Upper,* hybridization bands at 0.6 kilobase pair (*arrow*) of the total RNA (20 μg) of the fifth-instar 0-day *Bombyx* brain (*left*) and graded amounts of bombyxin RNA produced from λBb360 cDNA with T7 polymerase. To each sample of bombyxin RNA was added the posterior silk gland total RNA (20 μg) as carrier. Hybridization probe was multiprimed $^{32}$P-labeled λBb360 cDNA. *Lower,* standard curve for estimating the λBb360-type bombyxin mRNA amount. *Abscissa,* amount of synthetic bombyxin RNA; *ordinate,* densitometric values in arbitrary units for hybridization bands; *open circle,* datum point for the fifth-instar 0-day brain total RNA placed on the curve according to the densitometric value for hybridization band. λBb360-type bombyxin mRNA contained in this preparation is estimated as 80 pg from the corresponding point of *abscissa. B,* change in the amount of λBb360-type bombyxin mRNA in *Bombyx* brain during larval-pupal development. *Top,* Northern hybridization bands of brain total RNA (20 μg) from the animals various days after fourth- to fifth-instar ecdysis (*E*), wandering (*W*), and pupal ecdysis (*P*). Hybridization probe was $^{32}$P-nick-translated pBb360. *Middle,* the curve plotting the amounts of λBb360-type bombyxin mRNA/μg of brain total RNA for each developmental stage. Each datum point was determined as in *A. Bottom,* the above curve is redrawn on the per brain basis.

cDNA must have been detected. Thus, although the precise number of gene copies is unknown, it may be safe to conclude that more than 10 bombyxin gene copies exist.

No differences were found in the hybridization pattern between both sexes indicating that the bombyxin genes are not linked with either W or Z sex chromosome.

*Northern Hybridization: Developmental Change of Bombyxin mRNA Amount in Bombyx Brain*—The mRNA in *Bombyx* brain was quantified by Northern hybridization using λBb360 cDNA as probe. For constructing a standard curve for this quantification, we used the bombyxin RNA synthesized *in vitro* from λBb360 cDNA with T7 polymerase. As illustrated in Fig. 5*A* for the case of the fifth-instar 0-day *Bombyx* brain, 20 μg of brain total RNA is inferred to contain 80 pg of bombyxin mRNA. The amount of λBb360-type mRNA (only λBb360-type bombyxin mRNA must have been quantified by this procedure for the same reason as mentioned in the previous section) is thus 4 pg/μg of brain total RNA. This value is calculated to be $1.1 \times 10^7$ molecules/μg of brain total RNA, based on a 0.6-kilobase pair size of the bombyxin mRNA.

Developmental change in the λBb360-type mRNA amount was examined during larval-pupal development of *Bombyx* by using the above procedure. The λBb360-type mRNA amount per brain total RNA decreased gradually with the growth in the fifth-instar (Fig. 5*B*, *top* and *middle*). When this curve is redrawn in terms of the mRNA per brain, however, the titer remains essentially unchanged at 1–2 pg/brain (Fig. 5, *bottom*).

## DISCUSSION

The present study has identified the bombyxin gene actually expressed in the *Bombyx* brain among the multiple genes. The structure of preprobombyxin characterized here was shown to resemble that of preproinsulin in the domain organization of prepeptide/B-chain/C-peptide/A-chain the feature of which was also deduced from the analysis of genomic DNA containing a similar sequence (11). Thus, we presume the existence of a single-chain precursor molecule

for bombyxin that assures the insulin-like tertiary structure and posttranslational processing to generate a mature bombyxin. Recently, a cDNA encoding the "molluscan insulin-related peptide (MIP)" of *Lymnaea stangalis* has been characterized and the MIP gene has been shown to be expressed in the cerebral light-green cells that control the growth of *Lymnaea* (24). From these studies it is now unequivocally established that invertebrate nerve cells produce peptides that share a common ancestral molecule with vertebrate insulin family peptides, the concept that had been suggested earlier mainly from immunological and biological studies (25, 26) and first been proved directly by primary structure determination of naturally occurring bombyxin (6, 7).

Amino acid sequencing of native bombyxin molecules revealed the presence of six different molecular species which differed from one another by replacement of a few amino acid residues (6–10) and purification data suggested the presence of other bombyxin molecules yet to be sequenced. Furthermore, the bombyxin molecules deduced from the cDNA structure in the present study differed slightly in their sequence from those elucidated by peptide analysis. Thus, the presence of highly heterogeneous bombyxin molecules is now well established. However, it has not been known whether this heterogeneity results from the genetic variation associated with the use of millions of *Bombyx* heads for purification or is due to the presence of multiple bombyxin gene copies in the *Bombyx* genome. Southern hybridization studies reported here which revealed more than 10 copies of the bombyxin gene in the genomic DNA prepared from a single individual of *Bombyx* show that the latter is the case, although the former possibility might also hold true. In fact, our unpublished study on the cloning of bombyxin genes revealed the clustered localization of multiple bombyxin genes on a single DNA fragment of *Bombyx*.[2] The presence of multiple bombyxin gene copies is in sharp contrast to vertebrate insulin genes which exist in a single copy except for the case of rat genes that have two copies (21).

Four pairs of mid-dorsal brain neurosecretory cells of *Bombyx* have been immunohistochemically shown to contain bombyxin (5). Since λBb360-type bombyxin mRNA is contained in the fifth-instar 0-day *Bombyx* brain at a concentration of 4 pg or $1.1 \times 10^7$ molecules/μg of brain total RNA and the four pairs of bombyxin-producing cells occupy ~1/250 volume of a *Bombyx* brain, the concentration of λBb360-type mRNA in these cells is calculated to be $2.8 \times 10^9$ molecules/μg of total RNA, based on the assumption that the total RNA is distributed evenly throughout the brain. This bombyxin mRNA concentration is comparable to the abundance of fibroin mRNA (1.3–5.1% (w/w) or $1.4–5.5 \times 10^9$ molecules/μg of total RNA in the posterior silk gland), which is synthesized in a remarkably high rate during a limited period of growth (27, 28). The fibroin gene exists in a single copy and manifests a remarkably high transcriptional activity (29, 30). In the case of bombyxin, it seems probable that the immense accumulation of mRNA is supported, at least in part, by multiplication of the gene.

It was demonstrated that the amount of bombyxin in the *Bombyx* brain rose significantly at around pupation (31, 32). By contrast, the bombyxin mRNA hybridizable to λBb360 cDNA per brain did not change appreciably at that time. Furthermore, the amount of bombyxin per brain (0.1–3 ng, Ref. 33) is several thousand times more abundant on the molar basis than its mRNA. Taken together it seems probable that the translational and posttranslational processes are important in regulating the bombyxin titer rather than transcriptional process, although the definitive conclusion must await similar studies for other types of bombyxin mRNA.

[2] A. Kawakami, M. Iwami, H. Nagasawa, A. Suzuki, and H. Ishizaki, manuscript in preparation.

*Acknowledgments*—We are grateful to C.-C. Hui, T. Ozaki, and T. Nagata for technical advice. We also thank M. Sasaki, E. Suzuki, M. Nasu, M. Kato, K. Soma, and I. Kubo for their technical assistance. We are grateful to W. Hara, K. Ueno, T. Suzuki, S. Ohta, and K. Matsuno for valuable discussions.

REFERENCES

1. Ishizaki, H., and Suzuki, A. (1980) in *Neurohormonal Techniques in Insects* (Miller, T. A., ed) pp. 244–276, Springer Verlag, NY
2. Bollenbacher, W. E., and Granger, N. A. (1985) in *Comprehensive Insect Physiology, Biochemistry and Pharmacology* (Kerkut, G. A., and Gilbert, L. I., eds) Vol. 7, pp. 109–151, Pergamon, Oxford
3. Ishizaki, H., Mizoguchi, A., Fujishita, M., Suzuki, A., Moriya, I., O'oka, H., Kataoka, H., Isogai, A., Nagasawa, H., Tamura, S., and Suzuki, A. (1983) *Dev. Growth & Differ.* **25,** 593–600
4. Nagasawa, H., Kataoka, H., Hori, Y., Isogai, A., Tamura, S., Suzuki, A., Guo, F., Zhong, X-C., Mizoguchi, A., Fujishita, M., Takahashi, S. Y., Ohnishi, E., and Ishizaki, H. (1984) *Gen. Comp. Endocrinol.* **53,** 143–152
5. Mizoguchi, A., Ishizaki, H., Nagasawa, H., Kataoka, H., Isogai, A., Tamura, S., Suzuki, A., Fujino, M., and Kitada, C. (1987) *Mol. Cell. Endocrinol.* **51,** 227–235
6. Nagasawa, H., Kataoka, H., Isogai, A., Tamura, S., Suzuki, A., Ishizaki, H., Mizoguchi, A., Fujiwara, Y., and Suzuki, A. (1984) *Science* **226,** 1344–1345
7. Nagasawa, H., Kataoka, H., Isogai, A., Tamura, S., Suzuki, A., Mizoguchi, A., Fujiwara, Y., Suzuki, A., Takahashi, S. Y., and Ishizaki, H. (1986) *Proc. Natl. Acad. Sci. U. S. A.* **83,** 5840–5843
8. Nagasawa, H., Maruyama, K., Sato, B., Hietter, H., Kataoka, H., Isogai, A., Tamura, S., Ishizaki, H., Semba, T., and Suzuki, A. (1988) in *Peptide Chemistry 1987* (Shiba, T., and Sakakibara, S., eds) pp. 123–126, Protein Research Foundation, Osaka, Japan
9. Maruyama, K., Hietter, H., Nagasawa, H., Isogai, A., Tamura, S., Suzuki, A., and Ishizaki, H. (1988) *Agric. Biol. Chem.* **52,** 3035–3041
10. Jhoti, H., McLeod, A. N., Blundell, T. L., Ishizaki, H., Nagasawa, H., and Suzuki, A. (1987) *FEBS Lett.* **219,** 419–425
11. Iwami, M., Kawakami, A., Ishizaki, H., Takahashi, S. Y., Adachi, T., Suzuki, Y., Nagasawa, H., and Suzuki, A. (1989) *Dev. Growth & Differ.* **31,** 31–37
12. Huynh, T. V., Young, R. A., and Davis, R. W. (1984) in *DNA Cloning Techniques: A Practical Approach* (Glover, D., ed) pp. 49–78, IRL Press, Oxford
13. Sakai, N., Mori, S., Izumi, S., Haino-Fukushima, K., Ogura, T., Maekawa, H., and Tomino, S. (1988) *Biochim. Biophys. Acta* **949,** 224–232
14. Sanger, F., Nicklen, S., and Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. U. S. A.* **74,** 5463–5467
15. Kumar, V., Green, S., Staub, A., and Chambon, P. (1986) *EMBO J.* **5,** 2231–2236
16. Maniatis, T., Fritsch, E. F., and Sambrook, J. (1982) *Molecular Cloning, A Laboratory Manual,* Cold Spring Harbor Laboratory, Cold Spring Harbor, NY
17. Weeks, D. P., Beerman, N., and Griffith, O. M. (1986) *Anal. Biochem.* **152,** 376–385
18. Southern, E. M. (1975) *J. Mol. Biol.* **98,** 503–517
19. Turner, A. J. (1986) in *Essays in Biochemistry* (Marchall, R. D., and Tipton, K. F., eds) Vol. 22, pp. 69–119, Academic Press, London
20. Schwartz, T. W. (1986) *FEBS Lett.* **200,** 1–10
21. Lomedico, P., Rosenthal, N., Efstratiadis, A., Gilbert, W., Kolodner, R., and Tizard, R. (1979) *Cell* **18,** 545–558
22. Kyte, J., and Doolittle, R. F. (1982) *J. Mol. Biol.* **157,** 105–132
23. Bell, G. I., Swain, W. F., Pictet, R., Cordell, B., Goodman, H. M., and Rutter, W. J. (1979) *Nature* **282,** 525–527
24. Smit, A. B., Vreugdenhil, E., Ebberink, R. H. M., Geraerts, W. P. M., Klootwijk, J., and Joosse, J. (1988) *Nature* **331,** 535–538
25. Kramer, K. J. (1985) in *Comprehensive Insect Physiology, Biochemistry and Pharmacology* (Kerkut, G. A., and Gilbert, L. I., eds) Vol. 7, pp. 511–536, Pergamon, Oxford
26. Thorpe, A., and Duve, H. (1984) *Mol. Physiol.* **5,** 235–260
27. Ohta, S., Suzuki, Y., Takiya, S., and Suzuki, T. (1988) *Dev. Growth & Differ.* **30,** 293–299
28. Suzuki, Y., and Suzuki, E. (1974) *J. Mol. Biol.* **88,** 393–407
29. Suzuki, Y., Gage, L. P., and Brown, D. D. (1972) *J. Mol. Biol.* **70,** 637–649
30. Gage, L. P., and Manning, R. F. (1976) *J. Mol. Biol.* **101,** 327–348
31. Ishizaki, H. (1969) *Dev. Growth & Differ.* **11,** 1–7
32. Ishizaki, H., Suzuki, A., Moriya, I., Mizoguchi, A., Fujishita, M., O'oka, H., Kataoka, H., Isogai, A., Nagasawa, H., and Suzuki, A. (1983) *Dev. Growth & Differ.* **25,** 585–592
33. Ishizaki, H., and Suzuki, A. (1984) in *Biosynthesis, Metabolism and Mode of Action of Invertebrate Hormones* (Hoffman, J., and Porchet, M., eds) pp. 63–77, Springer-Verlag, Berlin, Heidelberg, Federal Republic of Germany

*Proc. Natl. Acad. Sci. USA*
Vol. 86, pp. 0000–0000, ●●●●●●●● 1989
Biochemistry

# Structure and organization of four clustered genes that encode bombyxin, an insulin-related brain secretory peptide of the silkmoth *Bombyx mori*

(Insect development/neurosecretory hormone/multi-gene family)

ATSUSHI KAWAKAMI*, MASAFUMI IWAMI*, HIROMICHI NAGASAWA†, AKINORI SUZUKI†, AND HIRONORI ISHIZAKI*‡

*Biological Institute, Faculty of Science, Nagoya University, Nagoya 464-01, Japan; and †Department of Agricultural Chemistry, Faculty of Agriculture, The University of Tokyo, Tokyo 113, Japan

*Communicated by Donald F. Steiner, May 23, 1989*

**ABSTRACT** **Four genes encoding bombyxin have been located in a 14-kilobase *Bombyx* genomic DNA segment. All of these genes encode preprobombyxin, the precursor molecule for bombyxin, with the domain organization of signal peptide/B chain/C peptide/A chain. Bombyxins are classified as family A or B according to their sequence homology. Two genes, each belonging to a different family, are closely apposed to form a pair with opposite orientation, presumably forming a regulatory unit for transcription. Genomic Southern blot hybridization suggested that there are many such gene pairs in the *Bombyx* genome. Differences between bombyxin genes and vertebrate insulin-family genes indicate that different mechanisms operate in the evolution of invertebrate and vertebrate insulin-family genes.**

Bombyxin, a 5-kDa brain secretory peptide of the silkmoth *Bombyx mori*, belongs to the insulin family (1, 2). Bombyxin has been purified by using its biological activity of stimulating the prothoracic glands to release ecdyson when applied to another moth, *Samia cynthia ricini* (3), but its physiological function in *Bombyx* is still obscure. Bombyxin is a heterogeneous group of molecular species that differ by only a few amino acid substitutions. Bombyxin molecules are heterodimers of A and B chains whose amino acid sequences are homologous with vertebrate insulin-family peptides (1, 4–6). The A and B chains of bombyxins are linked together by disulfide bonds, exactly as is insulin (7). We have cloned and characterized two distinct bombyxin cDNAs (8) and one bombyxin gene (9). Genomic Southern hybridization analyses using a bombyxin cDNA as probe indicated that multiple copies of bombyxin gene, as many as about 20, existed in the *Bombyx* genome. We now report the cloning and characterization of four bombyxin genes[§] that form a cluster in a 14-kilobase (kb) region of the *Bombyx* genome.

## MATERIALS AND METHODS

**Cloning and Sequencing of Bombyxin Genes.** Clone λ4k105, one of the four clones previously isolated (9), was used in this study. The nucleotide sequence was determined by the dideoxynucleotide chain-termination method (10) with overlapping deletion plasmids created by unidirectional digestion with exonuclease III and mung bean nuclease (11). Nucleotide sequence data were assembled and analyzed by using programs in DNASIS (Hitachi Software Engineering).

**Primer Extention.** Oligonucleotide primers were chemically synthesized. They are 5'-CGTGTGCACTCTTTGTGGCTGTTG-3' (family A primer; as is described in detail in *Results*, bombyxin genes are classified into families A and B according to their nucleotide sequence homology) and 5'-ACACATCAGACTGATCACGATTAC-3' (family B primer). The reaction was done with avian myeloblastosis virus reverse transcriptase (12) using 1.5 μg of *Bombyx* brain poly(A)$^+$ RNA from fifth-instar larvae.

**Probes for Northern and Southern Hybridizations.** As a probe for detecting family A members in Northern and Southern hybridization analyses, we used bombyxin gene A-1 (positions 505–1045 in figure 1 of ref. 9). For the detection of family B members, bombyxin gene B-1 (positions 370–996, see Fig. 2) was used as probe. The DNA fragments were nick-translated with [α-$^{32}$P]dCTP to give the specific activities of 1–3 × $10^8$ cpm/μg of DNA.

**RNA Preparation and Northern Blot Analysis.** The total RNA was "mini-prepared" by AGPC method (13) from various tissue sources of *Bombyx*. Aliquots (10 μg) of total RNA were electrophoresed on formaldehyde-denaturing 1.2% agarose gels (14) and transferred to nylon membrane filters with 20× SSC (3 M NaCl/0.3 M sodium citrate, pH 7.0). After prehybridization, the filters were hybridized at 37°C for 24 hr in 50% (vol/vol) formamide/5× SSC/5× Denhardt's solution (1× Denhardt's solution = 0.02% Ficoll/0.02% polyvinylpyrolidone/0.02% bovine serum albumin)/0.01% tRNA/0.01% denatured salmon sperm DNA/20 mM sodium phosphate, pH 6.5/0.1% SDS/10% (wt/vol) dextran sulfate and the probes at 1.5 × $10^6$ cpm/ml. The filters were washed in 2× SSC/0.1% SDS at 60°C for 3 hr followed by 0.2× SSC/0.1% SDS at 65°C for 1.5 hr.

**Genomic DNA Preparation and Southern Blot Analysis.** High molecular weight DNA was prepared from *Bombyx* silk glands of fifth-instar larvae as described (15). Aliquots (5 μg) of the genomic DNA were digested with restriction enzymes, separated by electrophoresis on 0.7% agarose gel, and alkali-transferred (16) to the nylon membrane filters. Hybridization was done at 60°C for 36 hr in 6× SSC/10× Denhardt's solution/20 mM sodium phosphate, pH 6.5/0.02% yeast tRNA/0.02% denatured salmon sperm DNA/0.5% SDS/10% dextran sulfate and the probes at 4–5 × $10^6$ cpm/ml. The filters were washed as for Northern blot hybridization.

## RESULTS

**Overall Organization of Four Bombyxin Genes in the *Bombyx* Genomic DNA Insert in Clone λ4K105.** Fig. 1 illustrates the restriction map of the *Bombyx* genomic DNA insert in a λ4K105 clone. This DNA hybridized with probes A and B at

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviation: MIP, molluscan insulin-related peptide.

‡To whom reprint requests should be addressed.

§The sequence reported in this paper has been deposited in the GenBank data base (accession no. ●●●●●●●●).

MZ6068, 26069, and 26070

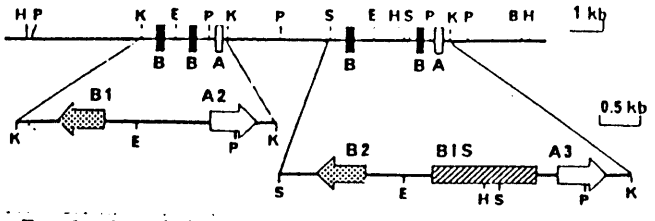



FIG. 1. Restriction map of a *Bombyx* genomic DNA segment in clone λ4K105. Fragments that hybridized to oligonucleotide probes A and B are indicated by open and solid bars, respectively. Sequenced regions are depicted in expanded scale to show four preprobombyxin genes. The open and stippled arrows represent family A (A-2 and A-3) and family B (B-1 and B-2) genes, respectively, and show the direction of transcription. BIS, a transposon-like inserted element; B, *Bam*HI; E, *Eco*RI; H, *Hin*dIII; K, *Kpn* I; P, *Pst* I; S, *Sal* I.

two and four sites, respectively. The hybridizing sites apparently formed two clusters and the nucleotide sequence was determined for these two regions (Fig. 2) to reveal four bombyxin genes (as shown in expanded scale in Fig. 1). As is discussed in the following section, bombyxin genes were classified into two families, termed families A and B. Two genes, each of which belongs to the different families, are paired (genes B-1/A-2 and B-2/A-3) with opposite orientation. Genes B-1 and A-2 are ≈1.1 kb apart and genes B-2 and A-3 are ≈2.2 kb apart. The wider space between B-2 and A-3 is due to the presence of an insertion sequence, designated BIS (≈1.2 kb). The failure of probe A to hybridize with genes B-1 and B-2, as shown in the restriction map, is accounted for by a low degree of the sequence matching between the probe and the A-chain portion of these genes.

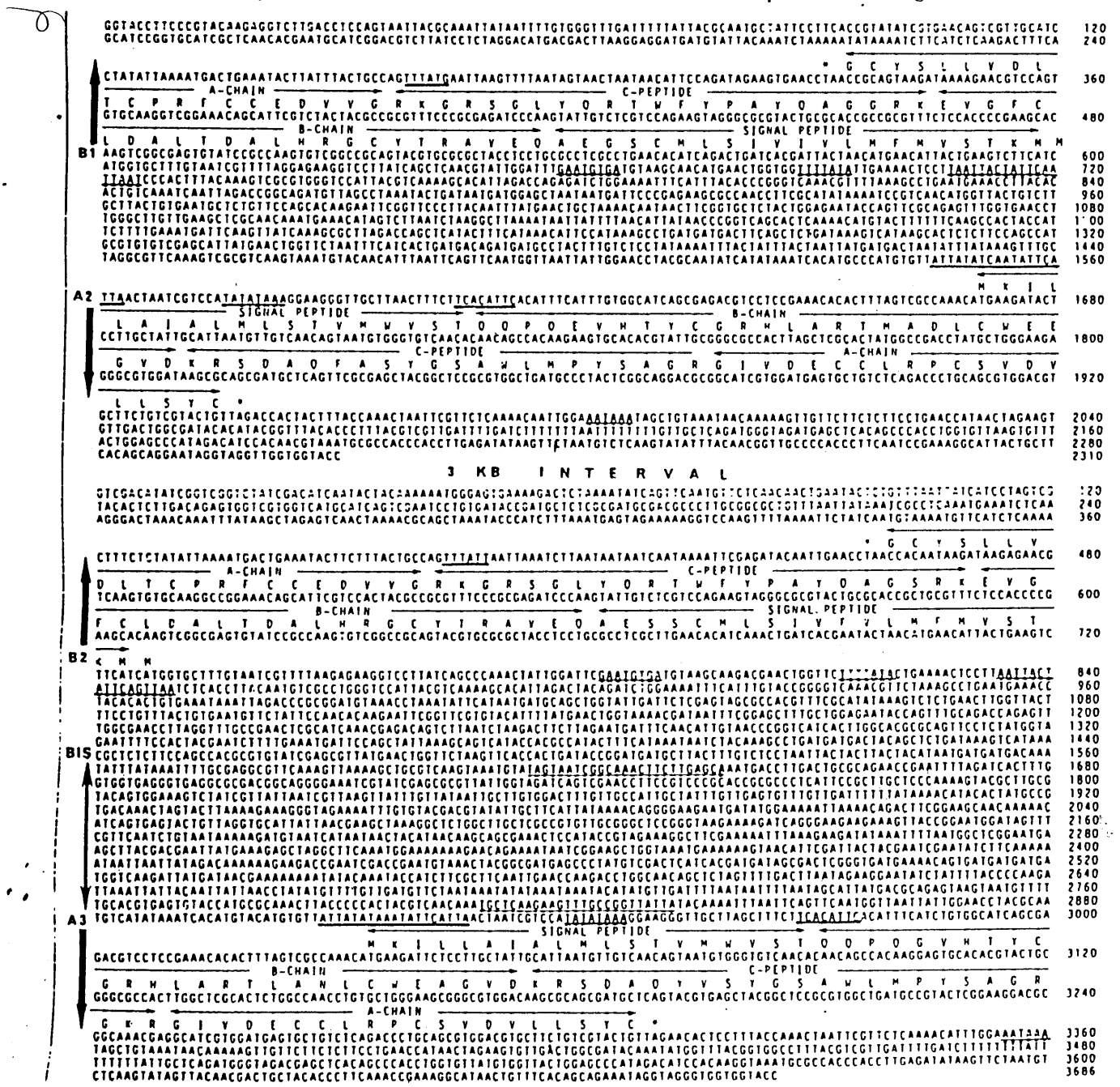



FIG. 2. Nucleotide and deduced amino acid sequences of the two regions of the λ4K105 *Bombyx* genomic DNA segment that carry four preprobombyxin genes. The sequences correspond to the regions depicted in expanded scale in Fig. 1. Arrows on the left cover the respective genes and show the direction of transcription. ~~Family B genes are shown as an antisense strand.~~ Stop codons are marked with asterisks. "TATA" boxes, polyadenylylation signals, cap consensus sequences and A+T-rich sequences are underlined. BIS is a transposon-like inserted sequence with terminal inverted repeats indicated by horizontal arrows below the sequence.

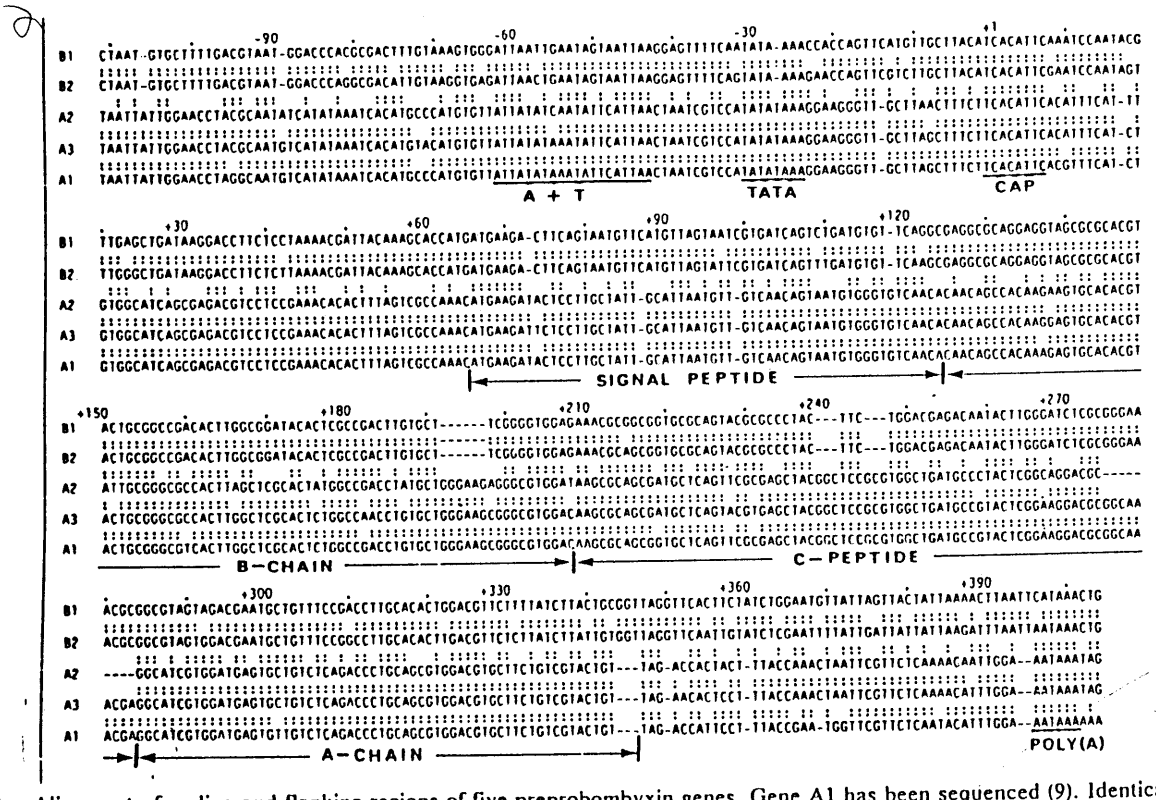



FIG. 3. Alignment of coding and flanking regions of five preprobombyxin genes. Gene A1 has been sequenced (9). Identical nucleotides between the neighboring sequences are indicated by colons.

**Structure of Bombyxin Genes and Their Flanking Regions and Classification of Bombyxin Genes into Families A and B.** Nucleotide sequences of the two regions containing the bombyxin genes are shown in Fig. 2. Each gene contains two or three candidates for the translation initiation codon but we postulate the initiation sites to be those indicated in Fig. 2 because of the presence of general features for a signal peptide. The open reading frame of all bombyxin genes apparently encodes a preprobombyxin, which is composed of four domains in the order signal peptide/B chain/C peptide/A chain. The structure of preprobomyxins suggests that they are posttranslationally modified to form mature bombyxins through the excision of the C peptide and the conversion of the glutamine at the N terminus of the B chain to pyroglutamate. These inferences have been made (8, 9) for the bombyxin gene and cDNAs characterized.

In Fig. 3 the coding and flanking sequences of the four genes reported here and of the characterized bombyxin gene (9) are aligned. The sequence homology is remarkably high (≈90%) between the upper two sequences and among the lower three sequences, whereas far less homology (≈50%) is seen between these two groups. Based on this sequence comparison, we tentatively classify the bombyxin genes into two families, families A and B. The characterized gene (9) belongs to family A and we now designate it gene A-1. The nucleotide sequence of λBb360 cDNA (8) completely matches the corresponding region of gene A-2. All of the genes lack introns.

A+T-rich, "TATA-box," transcription-start, and polyadenylylation-signal sequences are highly conserved throughout the five genes with respect to their sequences and positions (Fig. 3). The transcription start sites were determined by primer extension (data not shown). Besides these sites, no conserved sequences were found in the flanking regions including sequences outside of the area shown in Fig. 3.

When two gene pairs, B-1/A-2 and B-2/A-3, are compared by Harr plot homology analysis (●●●), the homology in the flanking regions extends for a considerable distance in both the 3′ and 5′ directions, except for the BIS insertion (Fig. 4). Thus gene pairs B-1/A-2 and B-2/A-3 form a highly homologous tandem repeat, suggesting that these gene pairs must have been generated by duplication of an original gene pair.

The BIS insertion has 25-base-pair inverted repeats at both ends (Fig. 2) that are flanked by 2-base-pair (TA) target-site duplications, indicating that this insertion may be a transposable element (17). A GenBank (R50.0) homology search (●●●●●●●●) failed to find sequence homology between BIS and any insertion or transposon sequence thus far identified.

**Amino Acid Sequence Comparisons of Preprobombyxins, Bombyxins, and Other Insulin-Family Peptides.** The sequences of preprobombyxins deduced from family A and B genes, a preprobombyxin deduced from the λBb204 cDNA (8), and native mature bombyxins (1, 4, 6) are aligned in Fig. 5 with the sequences of human insulin and the molluscan insulin-related peptide (MIP, ref. 19). The homology is high

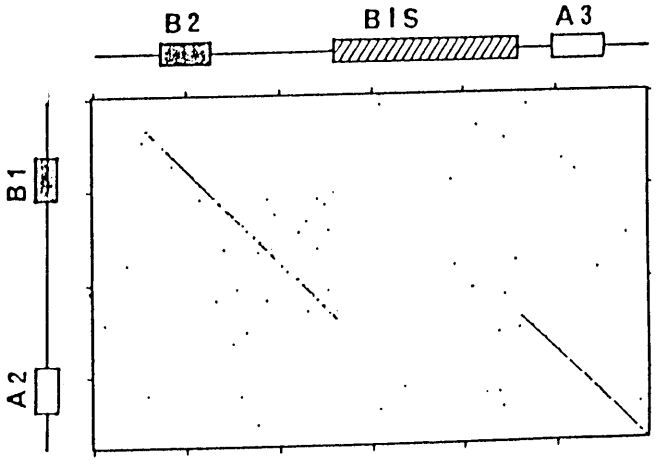



FIG. 4. Harr plot dot matrix for the two segments carrying gene pairs B-1/A-2 and B-2/A-3. The check-size/match-base number is 9/9.

DNAS 1 (V 3.0) of DNASIS (Hitac ... was used.

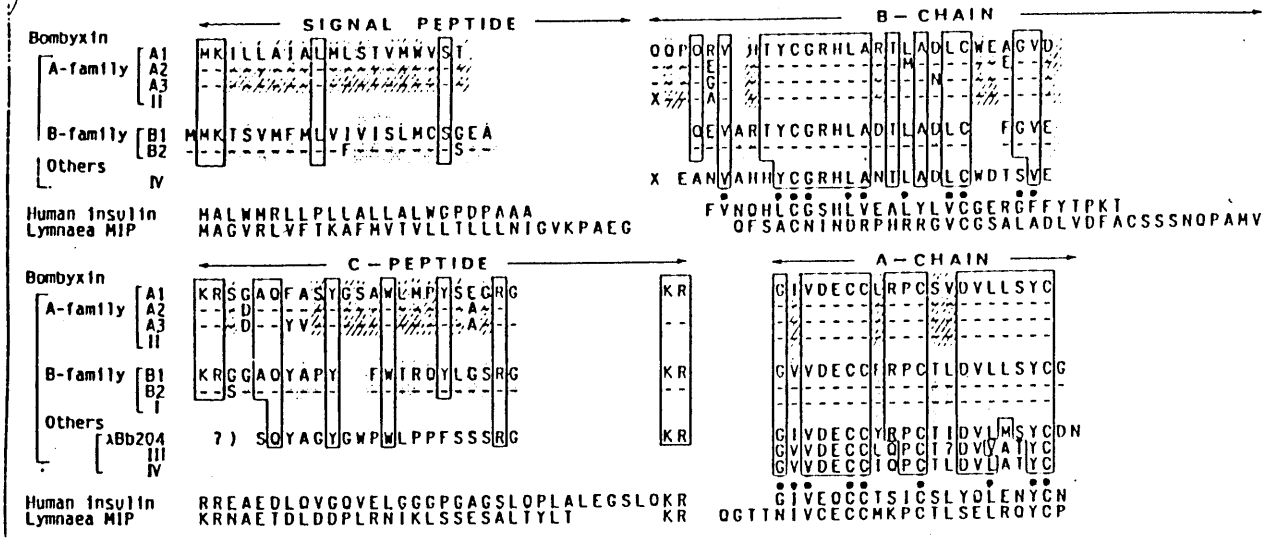



FIG. 5. Amino acid sequence comparison between preprobombyxins, bombyxins, and other insulin-family prepropeptides. Preprobombyxins A-2, A-3, B-1, and B-2 were deduced from the genes characterized in the present study. Preprobombyxin A-1 was deduced from the gene (9) and λBb204 is a part of the putative preprobombyxin deduced from the λBb204 cDNA (8). Bombyxin-I through -IV are mature peptides purified from *Bombyx* heads (1, 4, 6). Prepropeptides of human insulin (18) and molluscan insulin-related peptide (19) are also aligned. Gaps were introduced for maximum alignment. Homologous residues throughout bombyxin families A and B are boxed and the boxes are extended where identity was observed to other bombyxin molecules that do not belong to these families. Residues that are conserved only within bombyxin families A or B are indicated by stippled and hatched areas, respectively. The residues important for the formation of insulin tertiary structure (5), which are conserved or conservatively substituted in bombyxins, are indicated by the circles above the insulin sequence. X, pyroglutamate. The single-letter amino acid code was used.

within family A or family B but is low between families A and B. Bombyxin-II appears to belong to family A, whereas bombyxin-I seems to belong to family B, although these assignments are preliminary until the genes are characterized and the entire nucleotide sequences are compared. The amino acid sequences of bombyxin-III and -IV and of the putative partial preprobombyxin deduced from the λBb204 cDNA are quite different from other bombyxins and cannot be assigned to either family A or B, suggesting that still other bombyxin families remain to be defined. A very low degree of homology is observed between bombyxins and *Lymnaea* MIP. This large difference between insect and molluscan peptides suggests extensive evolutionary diversification of invertebrate insulin-related peptides.

When amino acid sequences of family A and B bombyxins were compared with the sequence of insulin, striking similarities were found between residues that contribute to the tertiary structure of insulin and corresponding residues in bombyxins, as was discussed (5) mainly for bombyxin-II. Thus, all residues responsible for the hydrophobic core formation are conserved or substituted by hydrophobic residues (A2, A3, A16, A19, B2, B6, B11, B12, B15, B18, and B24; taking the N termini of the A and B chains of insulin as position 1); A1, B8, and B23 are glycine, which contributes to the main chain conformation; all cysteines are conserved. Thus it seems highly probable that family A and B bombyxins have an insulin-like globular structure.

**Bombyxin Family A and B Genes Are Expressed in Brain.** To study the tissue-specific expression of family A and B bombyxin genes, Northern blot hybridization of RNA from various *Bombyx* tissues from day-0 fifth-instar larvae and freshly ecdysone-treated pupae was performed using genes A-1 and B-1 as probe. Preliminary hybridization experiments using these probes and subcloned family A and B bombyxin DNAs had shown that these probes hybridized specifically to their respective family DNAs under stringent conditions. RNA was isolated from subesophageal ganglion, fat body, silk gland, brain, malpighian tubule, ovary, and testis. Northern blots of these RNAs were made and hybridized to probes A and B. Brain was the only tissue examined that contained a transcript (≈6 kb) that hybridized to both probes, indicating that family A and B genes are specifically expressed only in the brain.

**Numerous Copies of Family A and B Genes Exist in the *Bombyx* Genome.** As shown in Fig. 6, multiple bands appeared when *Bombyx* genomic DNA digested with various restriction enzymes was hybridized with bombyxin genes A-1 and B-1 as probes, indicating that many copies of family A and B bombyxin genes exist in the *Bombyx* genome. The 1-kb *Hin*dIII fragment detected by the family A probe contained a copy of bombyxin gene A-1 and the 2.5-kb *Sal* I fragment revealed by the family B probe carried a copy of bombyxin gene B-1. By using the band intensity of these fragments, we estimated the number of copies of family A and B genes to be 12 and 10, respectively. When the corresponding restriction

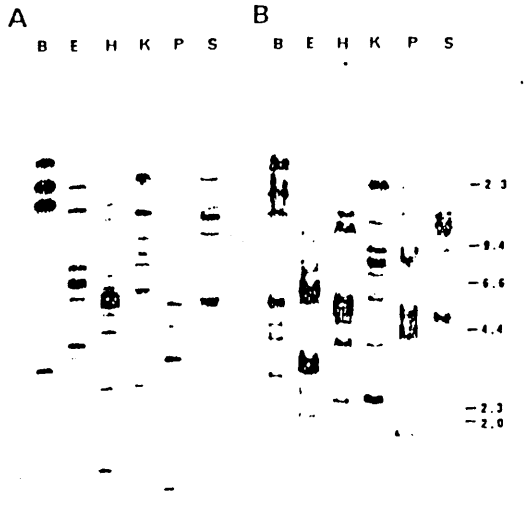



FIG. 6. Southern Genomic blot analysis. *Bombyx* genomic DNAs digested with various restriction enzymes as indicated by lane labels (abbreviations are as in Fig. 1) were probed by bombyxin genes A-2 (*A*) and B-1 (*B*).

lanes are compared, in many cases the bands that hybridized to the family A probe also hybridized to the family B probe. Since the family A and B probes did not cross-react, we suspect that the family A and B genes are localized close to one another, possibly forming pairs similar to the gene pairs B-1/A-2 and B-2/A-3. Furthermore, the fact that three intense bands were observed between 10 and 30 kb in the lanes containing *Bam*HI-digested DNA suggests that most of the bombyxin gene copies are clustered in three regions.

## DISCUSSION

The presence of multiple gene copies that encode bombyxin in the *Bombyx* genome has been suggested (8) by genomic Southern hybridization using a bombyxin cDNA as probe. The present study has verified this observation by demonstrating that a *Bombyx* DNA segment carries four bombyxin genes. The results of the Northern hybridization experiments and the structural features of all genes characterized in this study suggest that it is probable that all four genes are expressed. The deduced amino acid sequences of bombyxins differ somewhat from one another. Thus, we conclude that the bombyxin peptide polymorphism (1, 2, 6) results from heterogeneity at the genetic level (i.e., multiple copies in the *Bombyx* genome). Multiple bombyxin gene copies in the *Bombyx* genome and the lack of introns are in sharp contrast to vertebrate insulin-family genes that exist in one or two copies per haploid genome and have two or three introns (20). These differences suggest that evolutionary mechanisms for insulin-family genes in invertebrates and vertebrates differ significantly.

Remarkably, bombyxin genes are arranged in pairs in which the component genes belong to different families and are oriented in opposite directions. Genomic Southern hybridization experiments using genes A-1 and B-1 as probes have indicated that such gene pairs seem to occur frequently in the *Bombyx* genome. A similar gene-pair organization has been extensively studied for the moth chorion protein superfamily (21, 22). The chorion genes in each pair are coordinately expressed and the expression of various gene pairs is regulated in a developmental-stage-specific manner (23, 24). Coordinate expression of similar gene pairs has also been suggested for *Drosophila* genes that encode the salivary gland glue protein (25) and cuticular protein (26). Thus, it is probable that a bombyxin gene pair forms a unit for coordinate expression and various gene pairs are expressed in a temporally specific manner during the course of development.

Although bombyxin has a clear-cut prothoracicotropic function when applied to *Samia* (3), its function in *Bombyx* is still unknown. Interestingly, bombyxin has been detected in the *Bombyx* ovary and in embryos throughout embryonic development (27), in contrast to ●●●●●●●●, which is detectable only at the later stages of embryogenesis (28). It seems, therefore, likely that bombyxin might be involved in the control of oogenesis and embryogenesis, possibly through the regulation of cell proliferation and differentiation. Insulin and insulin-like growth factors control vertebrate cell proliferation and differentiation in a variety of tissues and developmental stages, including oocyte maturation (29) and embryogenesis (30). Detailed studies on the differential expression of bombyxin gene pairs might provide clues to the function of bombyxin.

We thank T. Takeda and M. Yamamoto (The Institute of Medical Science, The University of Tokyo) for preparing the oligonucleotide probes and primers. We also thank K. Soma and I. Kubo for technical assistance. This work was supported by grants from the Ministry of Education, Science and Culture of Japan, from The Ishida Foundation, and from The Mitsubishi Foundation.

1. Nagasawa, H., Kataoka, H., Isogai, A., Tamura, S., Suzuki, A., Mizoguchi, A., Fujiwara, Y., Suzuki, A., Takahashi, S. Y. & Ishizaki, H. (1986) *Proc. Natl. Acad. Sci. USA* 83, 5840–5843.
2. Ishizaki, H. & Suzuki, A. (1988) *Horm. Metab. Res.* 20, 426–429.
3. Ishizaki, H., Mizoguchi, A., Fujishita, M., O'oka, H., Kataoka, H., Isogai, A., Nagasawa, H., Tamura, S. & Suzuki, A. (1983) *Dev. Growth Differ.* 25, 593–600.
4. Nagasawa, H., Kataoka, H., Isogai, A., Tamura, S., Suzuki, A., Ishizaki, H., Mizuguchi, A., Fujiwara, Y. & Suzuki, A. (1984) *Science* 226, 1344–1345.
5. Jhoti, H., McLeod, A. L., Blundell, T. L., Ishizaki, H., Nagasawa, H. & Suzuki, A. (1987) *FEBS Lett.* 219, 419–425.
6. Maruyama, K., Hietter, H., Nagasawa, H., Isogai, A., Tamura, S., Suzuki, A. & Ishizaki, H. (1988) *Agric. Biol. Chem.* 52, 3035–3041.
7. Nagasawa, H., Maruyama, K., Sato, B., Hietter, H., Kataoka, H., Isogai, A., Tamura, S., Ishizaki, H., Senba, T. & Suzuki, A. (1988) in *Peptide Chemistry 1987*, eds. Senba, T. & Sakakibara, S. (Protein Res. Found., Osaka, Japan), pp. 123–126.
8. Adachi, T., Takiya, S., Suzuki, Y., Iwami, M., Kawakami, A., Takahashi, S. Y., Ishizaki, H., Nagasawa, H. & Suzuki, A. (1989) *J. Biol. Chem.* ~~in press.~~ 264, 7681–7685
9. Iwami, M., Kawakami, A., Ishizaki, H., Takahashi, S. Y., Adachi, T., Suzuki, Y., Nagasawa, H. & Suzuki, A. (1989) *Dev. Growth Differ.* 31, 31–37.
10. Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* 74, 5463–5467.
11. Henikoff, S. (1984) *Gene* 28, 351–359.
12. Calzone, F. J., Britten, R. J. & Davidson, E. H. (1987) *Methods Enzymol.* 152, 611–632.
13. Chomczynski, P. & Sacchi, N. (1987) *Anal. Biochem.* 162, 156–159.
14. Lehrach, H., Diamond, D., Wozney, J. M. & Boedtker, H. (1977) *Biochemistry* 16, 4743–4751.
15. Weeks, D. P., Beerman, N. & Griffith, O. M. (1986) *Anal. Biochem.* 152, 376–386.
16. Reed, K. C. & Mann, D. A. (1985) *Nucleic Acids Res.* 13, 7207–7221.
17. Rubin, G. M. (1983) in *Mobile Genetic Elements*, ed. Shapiro, J. A. (●●●●●●●●), pp. 329–361, 1431–1440. Academic Press, New Yo
18. Bell, G. I., Swain, W. F., Pictet, R., Cordell, B., Goodman, H. M. & Rutter, W. J. (1979) *Nature (London)* 282, 525–527.
19. Smit, A. B., Vreugdenhil, E., Ebberink, R. H. M., Geraerts, W. P. M., Klootwijk, J. & Joosse, J. (1988) *Nature (London)* 331, 535–538.
20. Steiner, D. F., Chen, S. J., Welsh, J. M. & Kwok, S. C. M. (1985) *Annu. Rev. Genet.* 19, 463–484.
21. Goldsmith, M. R. & Kafatos, F. C. (1984) *Annu. Rev. Genet.* 18, 443–487.
22. Kafatos, F. C., Mitsialis, S. A., Nguyen, H. T., Spoerel, N., Tsitilou, S. G. & Mazur, G. D. (1987) in *Development as an Evolutionary Process*, eds. Raff, R. A. & Raff, E. C. (Liss, New York), pp. 161–178.
23. Spoerel, N., Nguyen, H. T. & Kafatos, F. C. (1986) *J. Mol. Biol.* 190, 23–35.
24. Jones, C. W. & Kafatos, F. C. (1980) *Cell* 22, 855–867.
25. Mayerowitz, E. M. & Hogness, D. S. (1982) *Cell* 28, 165–176.
26. Snyder, M., Hunkapiller, M., Yuen, D., Silvert, D., Fristrom, J. & Davidson, N. (1982) *Cell* 29, 1027–1040.
27. Fugo, H., Hua, J., Nakajima, M., Nagasawa, H. & Suzuki, A. (1987) *J. Insect Physiol.* 33, 243–248.
28. Chen, J. H., Fugo, H., Nakajima, M., Nagasawa, H. & Suzuki, A. (1987) *J. Insect Physiol.* 33, 407–411.
29. Schorderet-Slatkine, M., El-Etr, & Baulieu, E.-E. (1979) *Science* 205, 1397–1399.
30. Froesch, E. R., Schmid, Chr., Schwander, J. & Zapf, J. (1985) *Annu. Rev. Physiol.* 47, 443–467.

the Bombyx prothoracicotropic hormone

to apper:
El-Etr, M., Schorderet-Slatkine, S. Baulieu, E.-E.

「遺　伝」　巻　　号（　年　月）

# 昆虫の変態をつかさどる遺伝子

岩　見　雅　史

昆虫の脱皮や変態の進行はホルモンにより支配されている．脳ホルモンは最も上位に位置するホルモンで，昆虫の成長，変態制御に中枢的な役割を果たしている．筆者らは最近，脳ホルモンの一つであるボンビキシンの遺伝子を単離し構造を決定した．遺伝子解析により得られた興味深い知見を紹介する．

## 1. 変態はホルモンにより支配されている

地球上には多種多様の昆虫が生息している．その数はヒト一人あたり3億匹ともいわれている．最も地球上を席巻している動物はヒト，そして昆虫である．このような昆虫の最も顕著な特徴は，その多くが変態を行なうことである．変態を伴う昆虫は，成長が劇的であるため，古くからヒトの目を引きつけてきた．

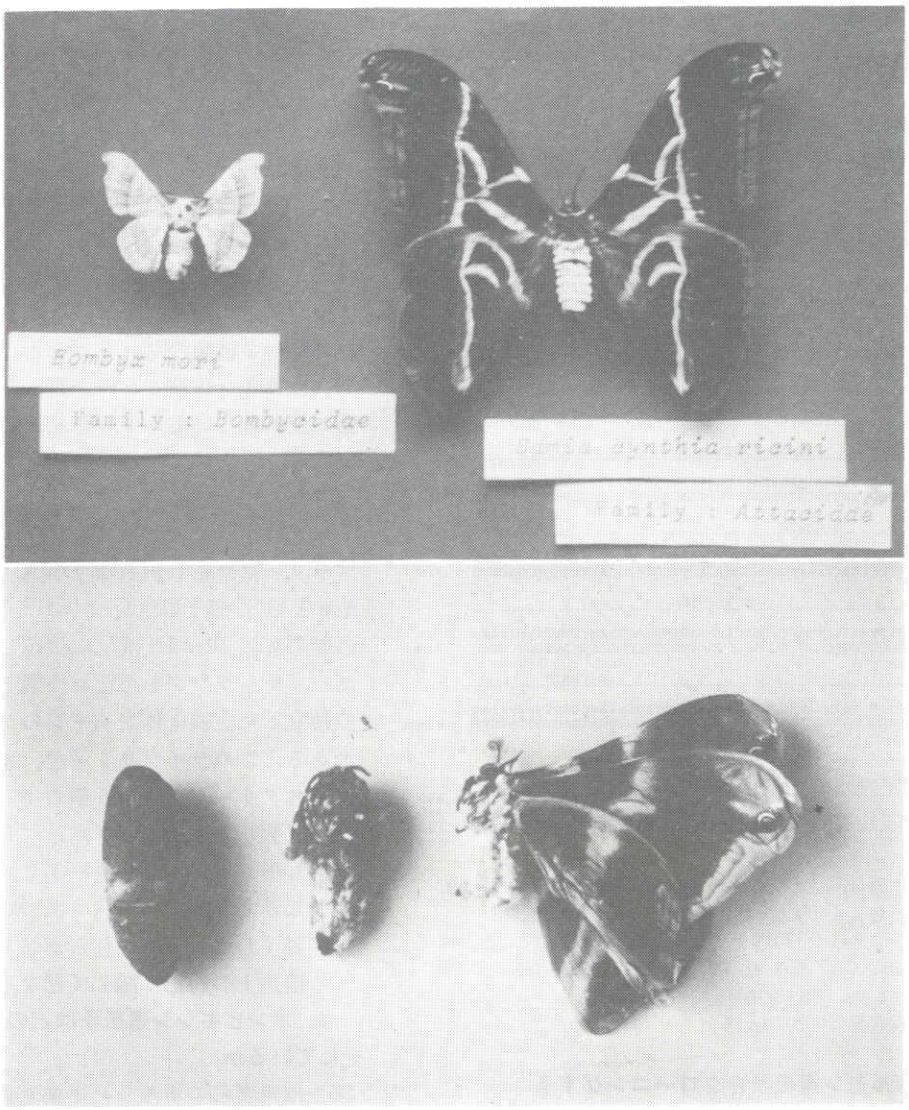
昆虫が科学の対象となったのは古いが，今世紀半ばになってようやく，脱皮や変態はホルモンによって支配されていることがわかった．そのホルモンとは脳ホルモン（前胸腺刺激脳ホルモン），エクジソン（前胸腺ホルモン），幼若ホルモン（アラタ体ホルモン）の三つのホルモンのことである．幼虫から幼虫への脱皮には幼若ホルモン存在下で脳ホルモンの刺激によりエクジソンが働くことが必要であり，幼虫から成虫への変態には幼若ホルモン不在下で脳ホルモンの刺激によりエクジソンが働くことが必要である．重要なことは，いずれの場合にも脳ホルモンが脱皮や変態の時期を支配していることである．

これら三つのホルモンのうち，ステロイドホルモンであるエクジソンは Butenandt と Karlson により精製，単離された後，Huber と Hoppe により1965年に，幼若ホルモンは Röller らにより1967年に構造が決定された．しかし，ペプチドホルモンである脳ホルモンは，ホルモンの生物検定法が難しいのに加えて，あまりにも微量しか存在しないことから，単離，構造決定は困難をきわめた（この単離，構造決定については石崎の解説[1,2]がある）．1984年に，ようやく200万倍の精製の後，カイコ脳ホルモンの一つであるボンビキシンの一次構造の一部が，そして1986年には全構造が明らかにされ，インスリンとの相同性が認められた[3]．

## 2. 脳ホルモンを単離し構造を決定する

ホルモンを単離し構造を決定しようとするとき，適切なホルモンの抽出源と活性の検定法を選ぶことが重要である．昆虫のなかでカイコは莫大な数（数百万頭単位）の供給が可能な唯一の昆虫であることから，カイコが脳ホルモンの抽出源として用いられた．

脳ホルモンの活性の検定には，まず蛹から脳を外科的に除去し発生を一時停止させる．発生を一時停止させられた除脳蛹に被検体を注射する．このとき，除脳蛹が発生を再開し変態するかどうかにより脳ホルモンの検定を行なう．しかし，この生物検定法に適した昆虫は必ずしも多くない．幸いにして，エリサンというヤママユガ科のガがほぼ理想的に検定動物としての条件を満たしていた．しかも，カイコの脳抽出液は種類の異なるエリサンの除脳蛹に対し変態を起こさせることができる（第1図）．脳ホルモンの抽出源としてカイコが，活性検定動物としてエリサンが用いられたことにより初めて脳ホルモンの単離，構造決定が可能となった．〔しかし近年，カイコの脳内には変態を起こさせうる脳ホルモンが2種類存在することがわかった．一つがエリサンやアゲハチョウに，もう一つがカイコに変態を起こさせうるものである[4]．前者をボンビキシン（Bombyxin），後者を PTTH（Prothoracicotropic hor-

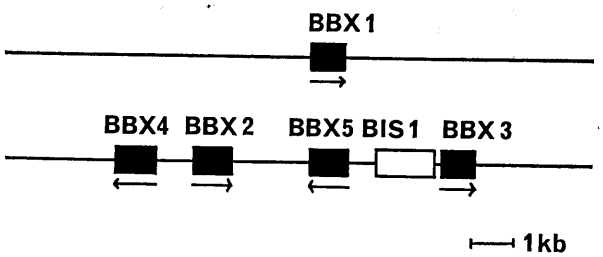



第1図　上：脳ホルモンの抽出源であるカイコ（カイコガ科：*Bombyx mori*）と活性検定動物であるエリサン（ヤママユガ科：*Samia cynthia ricini*）（原図石崎）．下：脳ホルモンの活性検定．脳を外科的に取り除いたエリサン除脳蛹（左）に脳ホルモン・ボンビキシンを注射すると成長を再開し，変態をとげる（中）．右のエリサン成虫に比べると翅を広げてはいないが「完全」な成虫である（原図石崎）．

mone）と呼んでいる．本稿では研究の進んでいる前者に話題をしぼって話を進める．］

このホルモン抽出，検定系を用いて，長沢らにより脳ホルモン・ボンビキシンのアミノ酸レベルでの構造が明らかにされた[3]．ボンビキシンはA鎖，B鎖二つの異なるペプチド鎖からなる二量体で分子量は約5キロダルトンである．ボンビキシンの一次構造，すなわちアミノ酸配列はインスリンと相同性をもっている．また，ボンビキシンには数種以上のアミノ酸配列の少し異なる分子が存在している．

ペプチド分析では出来上がった脳ホルモンの構造はわかっても，「どのように脳ホルモンが組み立てられるのか」また「どのような調節機構の下に生産されているのか」といった問題に答えることができない．脳ホルモンの遺伝子を単離し，DNAレベルで構造を解明し，発現調節のしくみを探ることが昆虫の変態を知るためのつぎの一歩と考えられる．




```
5'-AAGCTTGTGGAACCTCGCAACAAACGAGACCGTCTTAACCTAAGGCTTCAAAGAATGAT    59
   TTCAACGTTGTAACCCGGTCACCACTTGAAACAAGCAGTTTCTCTATGGTAAAATTTTCC   119
   ACTACCAATCTTTTGAAATGTTTCAAGCTATCAAAGCAATTGGACCACGCCATACTTTCA   179
   CAAATAAATCATAAAACATGATTACTACAGCTCTGATGAAGTCATAAACGCTCGCTTCCA   239
   GCCACGCGTGCATCGAGCGTTATGAACTGGTTCTAAGTTCACCACTGATACCGGATGATG   299
   CTTACTTTGTCTCCTAGTTACTACCTACTACATAATGAAGATGACAAATATTTATAAATT   359
   TTGCTAGGCGTTCAAAGTTAAAAGTTGCGTCAAGTAAATGTACAACATTTAATTCAGTTC  419
   ATTGGTTAATTATTGGAACCTAGGCAATGTCATATAAATCACATGCCCATGTGTTATTAT   479
   ATAAATATTCATTAACTAATCGTCCATATATAAAGGAAGGGTTGCTTAGCTTTCTTCACA   539
   TTCACGTTTCATCTGTGGCATCAGCGAGACGTCCTCCGAAACACACTTTAGTCGCCAAAC   599

                SIGNAL PEPTIDE
   -19          -10          -1
MetLysIleLeuLeuAlaIleAlaLeuMetLeuSerThrValMetTrpValSerThrGln
ATGAAGATACTCCTTGCTATTGCATTAATGTTGTCAACAGTAATGTGGGTGTCAACACAA   659

                B-CHAIN
            10                  20
GlnProGlnArgValHisThrTyrCysGlyArgHisLeuAlaArgThrLeuAlaAspLeu
CAGCCACAAAGAGTGCACACGTACTGCGGGCGTCACTTGGCTCGCACTCTGGCCGACCTG   719

                C-PEPTIDE
            30                  40
CysTrpGluAlaGlyValAspLysArgSerGlyAlaGlnPheAlaSerTyrGlySerAla
TGCTGGGAAGCGGGCGTGGACAAGCGCAGCGGTGCTCAGTTCGCGAGCTACGGCTCCGCG   779

                A-CHAIN
            50                  60
TrpLeuMetProTyrSerGluGlyArgGlyLysArgGlyIleValAspGluCysCysLeu
TGGCTGATGCCGTACTCGGAAGGACGCGGCAAACGAGGCATCGTGGATGAGTGTTGTCTC   839

            70
ArgProCysSerValAspValLeuLeuSerTyrCysAmb
AGACCCTGCAGCGTGGACGTGCTTCTGTCGTACTGTTAGACCATTCCTTTACCGAATGGT   899
TCGTTCTCAATACATTTGGAAATAAAAAAACCGTAAATAATAATAAAAAGTTCTTCTTTTC   959
CTGAACCACAATTGAAAGTGTTGACTGGAGATTTAGATACAGTTTTTTAGACATCATTTA  1019
GATGATTTTGATTGGCAAGCAAGCTT-3'                                  1045
```

第2図　上：脳ホルモン・ボンビキシン（BBX）の遺伝子地図．矢印はそれぞれ転写の向きを示している．kb＝キロ塩基対．下：ボンビキシン遺伝子（BBX 1）の塩基配列．影をつけた配列はTATA-box，CAT-box を示し，箱で囲った配列はキャップ配列を示している．下線部の配列は，ポリA付加シグナルおよび12塩基対の保存配列を示している．

### 3. 脳ホルモン遺伝子をクローニングする

一般にペプチドホルモンの遺伝子をクローニングする方法として三つの方法が利用されている．

1）ホルモンのアミノ酸配列を（部分的に）決定した後，アミノ酸配列から予想される遺伝子DNA の塩基配列を合成する．合成したDNAをプローブとして，遺伝子ライブラリーから目的のホルモンの遺伝子を単離する．

2）ホルモンに対する抗体を作製する．つぎに，cDNA ライブラリーを発現させ抗体と反応するクローンをスクリーニングすることにより目的の遺伝子を単離する．

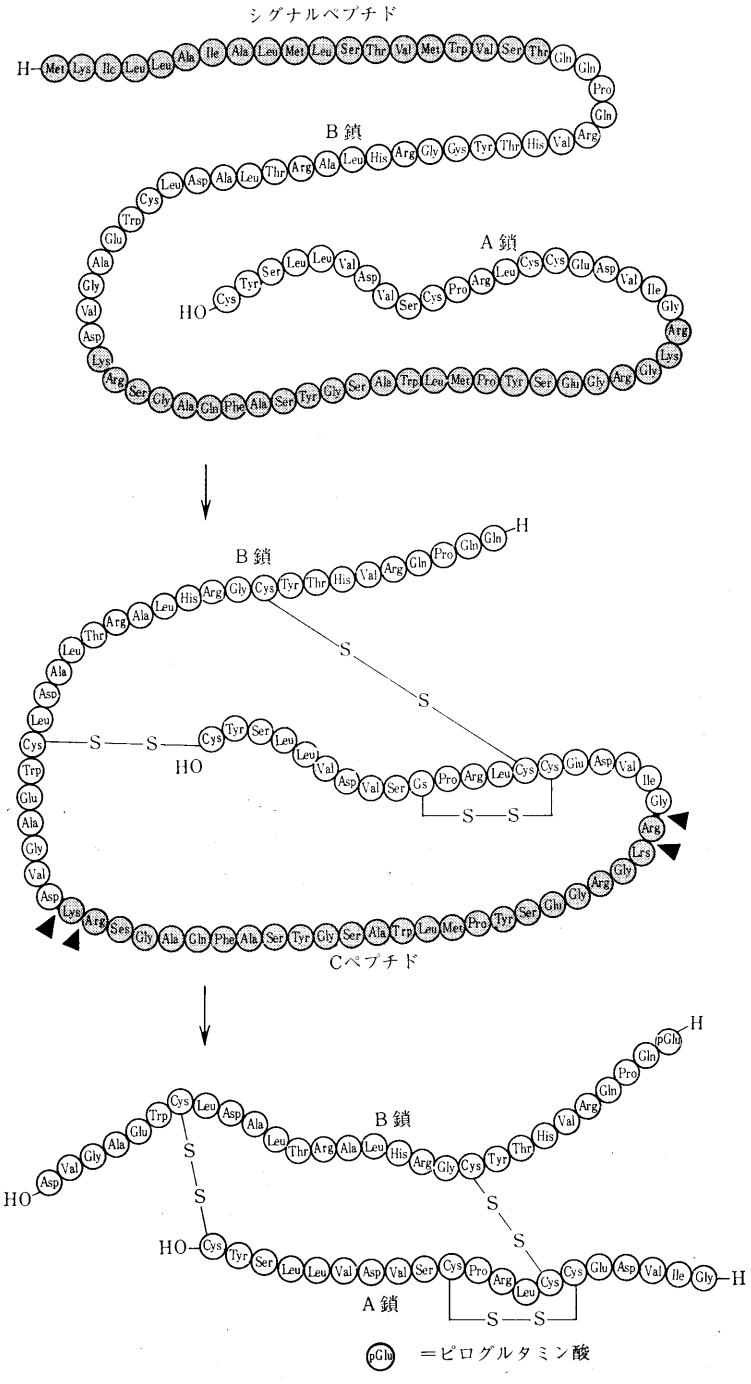
3）cDNA ライブラリーを発現させ，産物を生物検定によりスクリーニングし，目的の遺伝子を単離する．

筆者らは，すでに決定されたボンビキシンのアミノ酸配列を利用して，1）の方法により遺伝子の単離を試みた．ボンビキシンA鎖の4番目のアスパラギンから20番目のシステインまでの17個のアミノ酸に対する51塩基からなるDNA プローブAと，B鎖の8番目のスレオニンから24番目のグルタミン酸までの17個のアミノ酸に対する51塩基からなるプローブBの2種類のプローブを用いた．プローブAはカイコ体液タンパク質30 k[5]で最も頻繁に使われているコドンを利用して，プローブBはショウジョウバエのコドン利用度[6]を参考にして合成した．

スクリーニングに用いたカイコ遺伝子ライブラリーはカイコ絹糸腺 DNA を制限酵素 *Sau* 3 AI で部分消化し，ベクター λEMBL 3 の制限酵素 *Bam* HI 部位に接続し，試験管内でパッケージングを行ない作製した約12万個の独立した組換え体ファージのライブラリーである．このライブラリーを，まずプローブAでスクリーニングし6個のクローンを得た．6個のうち4個はプローブBとも反応しボンビキシン遺伝子をコードしていた．現在4個のうち2個のクローンについて遺伝子解析が終了したので，これらボンビキシン遺伝子の構造上の特色に関して話を進める．

### 4. ボンビキシン遺伝子はどのような構造をしているか

第2図上は筆者らの単離した組換え体ファージ中に存在するボンビキシン遺伝子（BBX）の地図である．BBX 1は1個の遺伝子が単独で存在している．BBX 2～BBX 5は対になった2個の遺伝子が重複した構造となっている．転写される向きは，BBX 2と BBX 3が右向き，BBX 4と BBX 5が左向きである．このような特徴的な構造はボンビキシンのほか昆虫ではコリオン，熱ショックタンパク質などの遺伝子でも見られる．これら遺伝子は，右向きのものと左向きのものとでは転写調節のしくみが異なっている[7,8]．ボンビキシン遺伝子では転写方向により別々の調節機構が存在するかどうかはまだわかっていない．また，ボンビキシン遺伝子の場合，BBX 3と BBX 5の間に約1200塩基対からなる挿入因子様

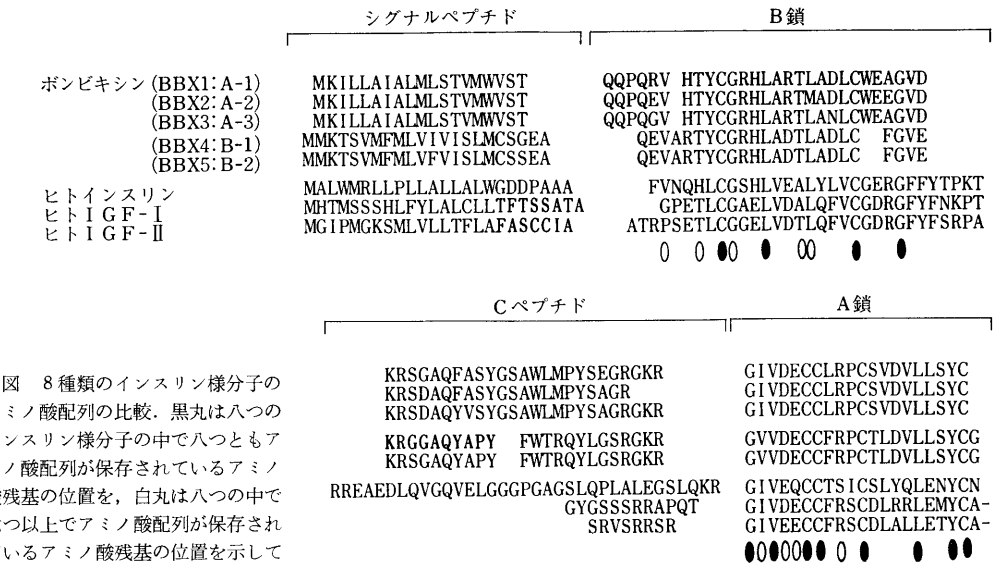



第3図　ボンビキシン分子の予想される生合成過程．▲はエンドペプチダーゼにより攻撃を受けると考えられる部位を示している．

配列（BIS 1）が存在している．この挿入因子様配列は両端に25塩基対の逆向き反復配列をもっている．

ボンビキシン遺伝子（BBX1）をもっと詳しく見てみよう（第2図下）．ボンビキシン遺伝子は276塩基対のコーディング領域（タンパク質に翻訳される領域）をもち，その遺伝子構成は5′側からシグナルペプチド（小胞体外へ分泌させるために必要なシグナル部位），B鎖，Cペプチド（A鎖とB鎖をつなぐペプチド），A鎖の順に並んでおり，ヒトをはじめとするインスリンとまったく同じ遺伝子構成である（第5図参照）．この遺伝子構成から，ボンビキシンはインスリンと同様な経路で生合成されると考えられる．すなわち，ボンビキシンはまず92アミノ酸のプレプロペプチドとして合成される．同時に，シグナルペプチドが切り離される．この時点でインスリンと似た立体構造が形成され，ジスルフィド結合が架けられた後，エンドペプチダーゼによりCペプチドが切りとられたボンビキシン分子ができあがる（第3図）．

これまでに五つのボンビキシン遺伝子の構造が明らかになっている．五つのプレプロボンビキシンのDNA塩基配列から予想されるアミノ酸配列を，ヒトプレプロインスリン，プレプロIGF（インスリン様成長因子）-Ⅰ，プレプロIGF-Ⅱのアミノ酸配列とともに比較したのが第4図である．五つのプレプロボンビキシンはアミノ酸配列の相同性から二つのグループ（Aと




第4図　8種類のインスリン様分子のアミノ酸配列の比較．黒丸は八つのインスリン様分子の中で八つともアミノ酸配列が保存されているアミノ酸残基の位置を，白丸は八つの中で六つ以上でアミノ酸配列が保存されているアミノ酸残基の位置を示している．

**B**）に分けることができる．BBX 1，BBX 2，BBX 3 は**A**グループ（順に A-1，A-2，A-3）に，BBX 4 と BBX 5 は**B**グループ（順に B-1，B-2）に分類されうる．各グループ内では，9割程度アミノ酸配列が保存されているが，グループ間では6割程度しか保存されていない．

第4図に掲げた八つのインスリン様分子の中で，アミノ酸配列が保存されている箇所を注意深く見てみると，これら分子が立体構造を形成するとき重要な役割を果たすアミノ酸残基，すなわち，

1）ジスルフィド結合を形成するアミノ酸残基

2）グリシン残基

3）立体構造の核となる疎水性アミノ酸残基

がよく保存されている[9]．

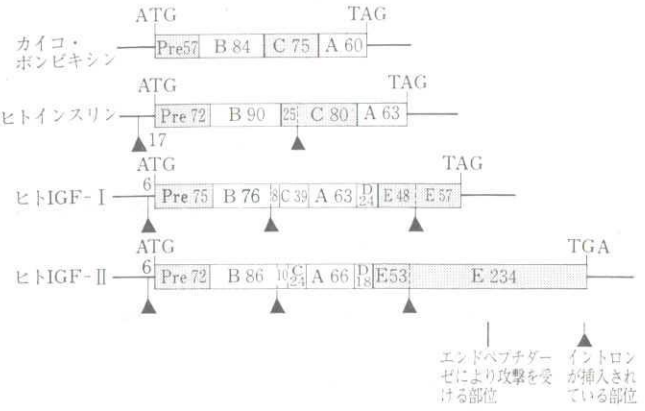
1）は骨組みの桟(さん)もしくは梁(はり)を形づくるアミノ酸残基で Cys (B 10)〔B鎖10番目のシステイン残基の意，以下同様〕，Cys (B 22)，Cys (A 6)，Cys(A 7)，Cys (A 11)，Cys (A 20)．2）は立体的に自由度が大きいため骨組みの曲りかどを形づくるアミノ酸残基で，Gly (B 11)，Gly(B 26)，Gly (A 1)．3）は立体構造の核，すなわちしんとなる疎水性アミノ酸残基で，Leu (B 14)，Leu (B 18)，Ile (A 2)，Val (A 3)，Leu (A 16)，Tyr (A 19)．

A鎖，B鎖以外の特徴を簡単に述べる．シグナルペプチドは，八つの分子の中で，アミノ酸配列は保存されていない．しかし，その長さ，そして疎水性のアミノ酸残基で構成されている点では八つの間で共通している．プレプロボンビキシンのCペプチドをプレプロインスリンのCペプチドと比較してみると，プロペプチドが適切な立体構造を形づくるために必要な長さと極性をもっている[10,11]．

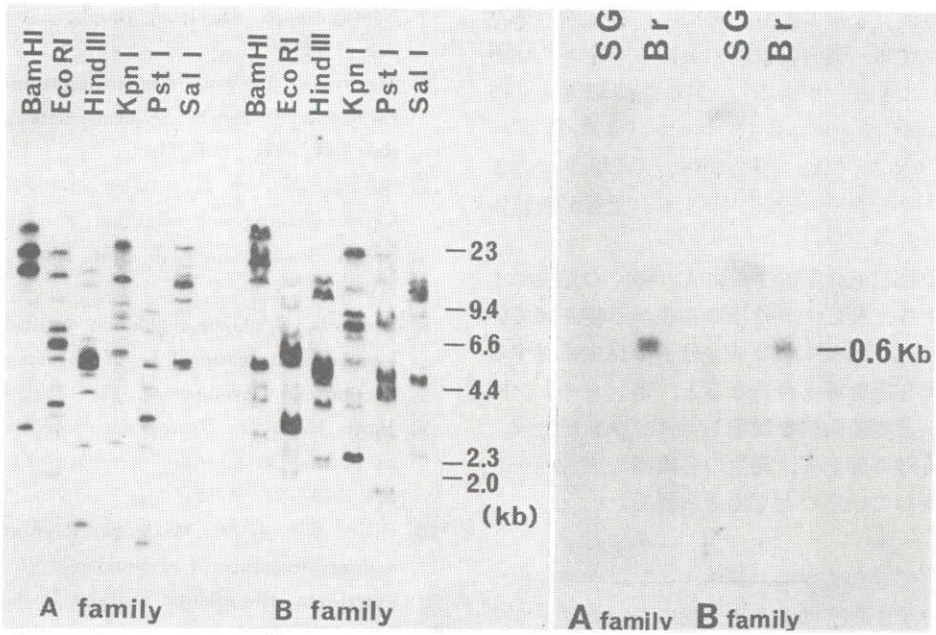
これまでに10種以上の動物からインスリン族の遺伝子が単離されている．これらの遺伝子はすべてコーディング領域の 5′ 側上流に一カ所，そしてラットインスリン遺伝子1を除いたすべての遺伝子のコーディング領域に一カ所もしくは二カ所イントロンが存在している[11]（第5図）．しかし，ボンビキシン遺伝子には，これまでに塩基配列を決定した五つの遺伝子すべてにイントロンが存在しない．カイコインスリン族遺伝子は進化の過程でイントロンを失ったのだろうか．カイコのインスリン族遺伝子にイントロンがないという事実は，インスリン分子の進化を解明するうえで興味深い問題を投げかけている．

## 5. ボンビキシン遺伝子のコピー数と脳での発現

ヒトやネズミのインスリン遺伝子はゲノム中に1コピーもしくは2コピーしか存在していない．ところが，ボンビキシン遺伝子をクローニングしてみると，カイコゲノム中のボンビキシン遺伝子のコピー数はどうも1コピーや2コピーではすまないようだ．実際に存在するコピー数を調べるため，ボンビ




第5図　ボンビキシン，ヒトインスリン，ヒトIGF-Ⅰ，ヒトIGF-Ⅱ遺伝子の遺伝子構成および産物の模式図．影をつけた部分は生合成の過程で切りとられるペプチドを示している．数字はおのおのの部分をコードしているDNAの塩基数を，Pre，B，C，A，D，Eは順に，シグナルペプチド，B鎖，Cペプチド，A鎖，Dペプチド，Eペプチドを，ATGは翻訳開始位置を，TAGおよびTGAは翻訳停止位置を示している．




第6図　左：サザンハイブリダイゼイション．上にカイコゲノムDNAを消化した制限酵素を示している．プローブはAグループ（**A family**）またはBグループ（**B family**）に特異的なものを用いた．使用したゲノムDNAはレーンあたり5 $\mu$g である．右：ノーザンハイブリダイゼイション．上にポリA（+）RNAの抽出源（SG：絹糸腺，Br：脳）を示している．プローブは左図に同じ．使用したポリA（+）RNAはレーンあたり2.5 $\mu$g である．Kb：キロ塩基対．

キシン遺伝子をプローブに使ったハイブリダイゼイションを行なった（第6図左）．A，B各グループのプローブとも制限酵素*Eco*RⅠや*Hin*dⅢで消化したゲノムDNA中に，10本以上のバンドが見られる．すなわち，ボンビキシン遺伝子はカイコゲノム中に少なくとも20コピー存在していることになる（プローブはA，B各グループに特異的である）．このことは，ペプチド分析により得られた「ボンビキシンはどうして数種以上のアミノ酸配列の少し異なる分子を含んでいるのか」という疑問に答えを与えるものである．

単クローン抗体を使った免疫組織化学的手法により，ボンビキシンを生産している神経分泌細胞（通常の神経細胞としての機能とホルモン分泌機能を兼ね備えた神経細胞）はカイコ脳の脳間部に4対存在している[12]．よってボンビキシンmRNAは脳に極在していることが期待される．実際，A，B各グループのプローブを用いたハイブリダイゼイションにより，予想通りボンビキシンmRNAは各グループとも脳に極在していることがわかった（第6図右）．

## 6. 脳ホルモン・ボンビキシンの機能は何か

インスリン様分子が昆虫にも存在していることから，インスリンは進化の過程でよく保存されており，発生や成長に重要な役割を果たしていると考えられる．脊椎(せきつい)動物では，インスリンはグリコーゲン，脂質，アミノ酸，タンパク質の合成に関与している．加えて，多くの培養細胞にマイトジェニック（DNA 合成をさかんにし成長を促す）効果をもたらす．IGF-I と IGF-II もマイトジェニック効果に関与している．

ショウジョウバエでは，脊椎動物のインスリンが成虫原基の成長を促したり，多くの培養細胞の増殖を進めたりする（無脊椎動物におけるインスリンの作用に関して総説[13]がある）．さらに，スズメガの一種（*Agrius cinglulata*）やオオカバマダラ（*Danaus plexippus*）では，脊椎動物のインスリンにより，トリグリセリド，ジグリセリド，脂質が脂肪体から放出される．

一方，免疫組織化学的手法により多くの昆虫にインスリン抗体と反応する分子の存在が示唆されている．さらに，これらインスリン分子の昆虫に対する多様な生理作用を考え合わせると，ボンビキシンはエリサンなどの前胸腺を刺激しエクジソンを合成，放出させるのに加えて，発生の各段階で成長因子としても機能していることも考えられる．

*　　*　　*

以上述べてきたように，脳ホルモンを含め昆虫ホルモンの分子生物学はようやく緒についたばかりである．前胸腺を刺激し，変態を起こさせる脳ホルモンの一つのボンビキシンが脊椎動物のインスリンと似ていることは脳—消化管系ホルモンの機能や進化に新たな問題を提出するものである．今後，クローニングされたボンビキシン遺伝子を使って，日周期，温度変化などにより左右される変態がどのような調節の下に制御されているかを明らかにすることが，つぎの研究の大きな目標の一つであろう．

本稿をまとめるにあたり指導ならびに助言を賜った石崎宏矩教授に，また大学院生川上厚志氏の協力に感謝する．

## 文　献

1) 石崎宏矩：変態ホルモン—脳ホルモン研究小史—．遺伝，**32** (3)，8—14 (1978).
2) 石崎宏矩：昆虫の変態を支配するホルモン—ついに構造を決定．科学朝日，**46** (8)，124—129 (1986).
3) Nagasawa, H. *et al.*: Amino acid sequence of a prothoracicotropic hormone of the silkworm *Bombyx mori*. Proc. Natl. Acad. Sci. U.S.A., **83**, 5840—5843 (1986).
4) Ishizaki, H. *et al.*: Species specificity of the insect prothoracicotropic hormone (PTTH): the presence of *Bombyx*- and *Samia*- specific PTTHs in the brain of *Bombyx mori*. Develop. Growth and Differ., **25**, 593—600 (1983).
5) Sakai, N. *et al.*: Structures and expression of mRNAs coding for major plasma proteins of *Bombyx mori*. Biochem. Biophys. Acta., **949**, 224-232 (1988).
6) Maruyama, T. *et al.*: Codon usage tabulated from the GenBank genetic sequence data. Nucl. Acids Res., **14**, r151—r197 (1986).
7) Spoerel, N., H. T. Nguyen and F.C. Kafatos: Gene regulation and evolution in the chorion locus of *Bombyx mori*. J. Mol. Biol., **190**, 23—35 (1986).
8) Richards, G.: Gene regulation during insect development. International J. of Invertebrate Reproduction and Development, **12**, 115—144 (1987).
9) Jhoti, H. *et al.*: Prothoracicotropic hormone has an insulin-like tertiary structure. FEBS Letters, **219**, 419—425 (1987).
10) Chan, S.J. *et al.*: Messenger RNA sequence and primary structure of preproinsulin in a primitive vertebrate, the atlantic hagfish. J. Biol. Chem., **256**, 7595—7602 (1981).
11) Steiner, D.F. *et al.*: Structure and evolution of the insulin gene. Ann. Rev. Genet., **19**, 463—484 (1985).
12) Mizoguchi, A. *et al.*: A monoclonal antibody against a synthetic fragment of bombyxin (4K-prothoracicotropic hormone) from the silkmoth, *Bombyx mori*: characterization and immonohistochemistry. Mol. Cell. Endocr., **51**, 227—235 (1987).
13) Kramer, K.J.: Vertebrate hormones in insects. Comprehensive insect physiology, biochemistry and pharmacology, **7**, 511—536, Pergamon Press, Oxford (1985).

（**いわみ まさふみ**，名古屋大学 理学部 生物学教室）