

論文 / 著書情報
Article / Book Information

題目(和文)	古細菌のFタイプATPase遺伝子の発見
Title(English)	
著者(和文)	鷲見正人
Author(English)	
出典(和文)	学位:博士(理学), 学位授与機関:東京工業大学, 報告番号:甲第3890号, 授与年月日:1998年5月31日, 学位の種別:課程博士, 審査員:
Citation(English)	Degree:Doctor (Science), Conferring organization: Tokyo Institute of Technology, Report number:甲第3890号, Conferred date:1998/5/31, Degree Type:Course doctor, Examiner:
学位種別(和文)	博士論文
Type(English)	Doctoral Thesis

目次

第1章 序論	1
第2章 PCR法によるバクテリアのF/VタイプATPase遺伝子の確認	6
第3章 古細菌からのFタイプATPase遺伝子の単離と塩基配列決定	15
第4章 古細菌におけるFタイプATPase遺伝子の存在とバクテリア進化との関係	26
引用文献	33

付録A *M. barkeri* FタイプATPase遺伝子の全塩基配列と予想されるアミノ酸配列

付録B *M. barkeri* FタイプATPase遺伝子産物と既知のFタイプ構成構成成分との相同性

付録C VおよびFタイプ、触媒および非触媒サブユニットのアライメント

第1章 序論

1977年にWoeseらによりはじめて提唱され、以後、遺伝子解析技術の進歩とそれに伴う情報の蓄積により定着した古細菌という生物群には、当初、真正細菌(Fタイプ)とは異なる、真核生物の空胞系にあるものと同じ種類のATPase(Vタイプ)のみが存在すると信じられていた。この状況から導き出された、古細菌と真正細菌がそれぞれ固有のタイプのATPaseを進化の過程で獲得したという説は魅力的であったが、真正細菌よりVタイプが見つかり修正の必要をせまられる。本研究は、F/Vタイプが真正細菌/古細菌の区別なく広範に分布する可能性を裏付ける情報を提供し、特にまだ報告されていない古細菌におけるFタイプの存在を示唆する情報を得るために取り組まれた。

1.1 F_oF₁-ATPase

生命体の能動的活動にはATPが必要である。このATPのほとんどは、電子伝達系の酵素群などを用いて膜の内外に生じさせた電気化学的ポテンシャル差を利用し、ADPとリン酸より合成される。ATP合成をつかさどる細胞内小器官としては、ミトコンドリアが有名だが、その内膜にありATP合成の役割を演じているのがF_oF₁-ATP合成酵素である。この酵素の反応は可逆で、ATPを水解することで膜を隔ててプロトンを輸送することが可能であること、F_oから遊離したF₁部分のみでATP水解能を持ち、これが多くの研究者の標品となっていることから、本酵素はATP分解酵素を意味するATPaseとも呼ばれる。

F_oF₁-ATPaseは、膜に埋め込まれた疎水性ペプチドよりなるプロトンチャンネルのF_o部分と、ATP合成/水解をになう触媒部位を含む大きな水可溶性部分のF₁より構成される、比較的大きく複雑で特徴的な構造を持つ酵素である[Boyer, 1990; Weber & Senior, 1997]。ミトコンドリア[Walker *et al.*, 1985]以外にも、やや仕組みが単純化された同種の酵素が、同じ真核生物である植物の葉緑体[Graber, 1987]、原核生物でも大腸菌をはじめとする様々な真正細菌[Dunn and Futai, 1980]より見つかっている。

各種生物より単離された同酵素のサブユニット組成を比較することで、この酵素の基本的構成成分が分かっている[Foster & Fillingame, 1982; Yoshida *et al.*, 1979]。F_o部分についてはa、b、cの3種類、F₁部分についてはα、β、γ、δ、εの5種類のポリペプチドよりなる。F₁部分は、交互にそれぞれ3つずつ並ぶ50kDa程度の大きなα、βというサブユニットに基質結合部位があり、そのうちβがATP合成/水解活性を持つことが知られている[Boyer, 1987; Hisabori *et al.*, 1986; Dancna *et al.*, 1986]（図01）。また、この大きな2つの主要サブユニットはその一次構造において、お互いに25%程度の相同性を持ち、かつて遺伝子重複により分化したものであることが示唆されている[Gorgarten *et al.*, 1989]。さらに、α、βそれぞれが種が変わっても50%程度の高い相同性を保つ。

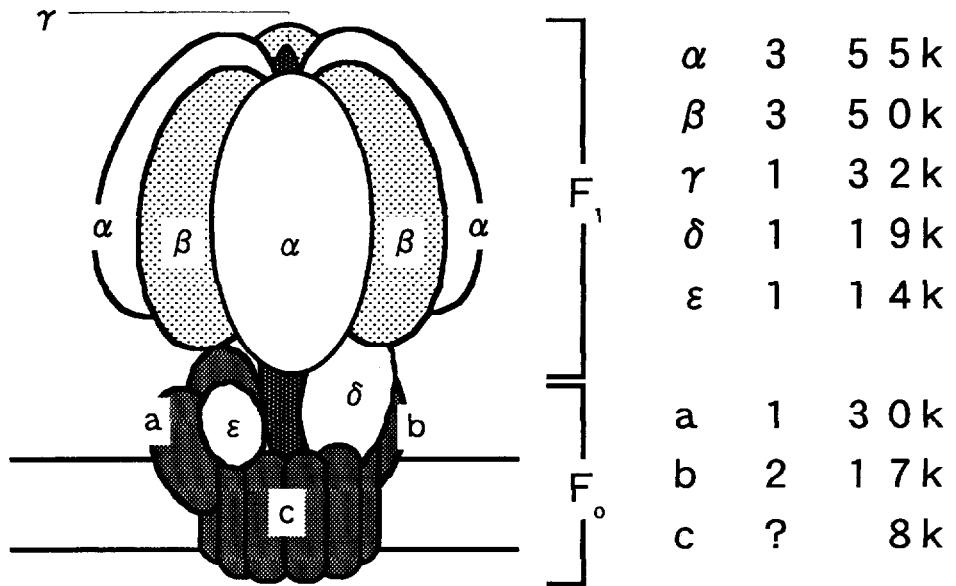


図01) F_0F_1 -ATPase の構造を示す略図。薄いグレイが触媒サブユニットの β で、濃いグレイが α , β の中心を貫く γ 。中庸のグレイは F_0 の構成成分を示す。右手の表は左からサブユニット名、酵素 1 分子に含まれる数、大腸菌におけるサブユニットの分子量である。c サブユニットの数はまだ決められていないが 10 から 12 と言われている。

1.2 F_0F_1 -ATPase の普遍的分布と古細菌における例外

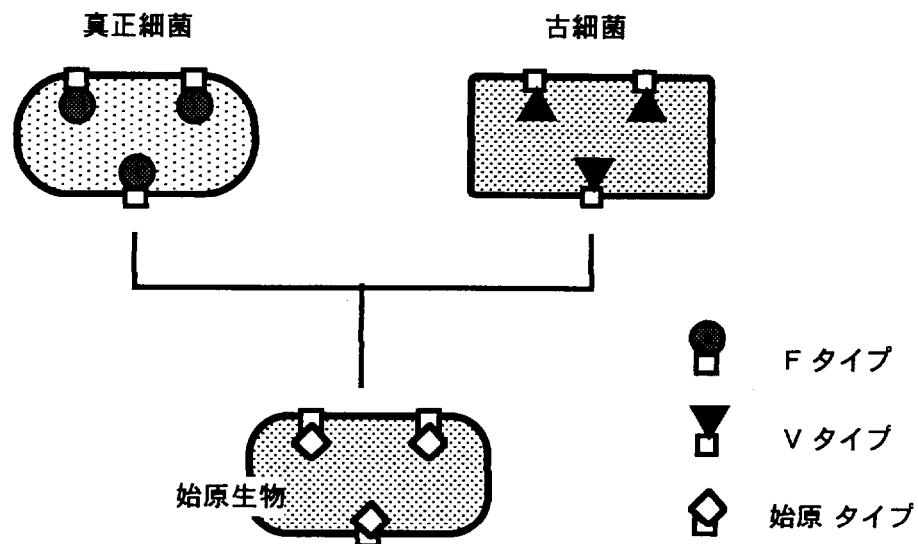
ATP の合成（もしくは逆反応による膜電位の生成）という機能の重要性や、バクテリアから高等生物までという幅広い分布と種間での保存性の高さより、当初は ATP 合成能を放棄した一部の寄生的細胞を除き、すべての生物が F_0F_1 -ATPase を持っていると考えられていた。そのため、1977 年 Woese らが、16S rRNA の類似性にもとづき、真核生物と真正細菌とは別に第 3 の生命分類群として提唱した古細菌[Woese et al., 1987]にも、 F_0F_1 -ATPase が存在するものであると信じられていた。

しかし、1980 年代、古細菌に属する好塩菌[Ihara et al., 1991]、メタン産生菌[Inatomi et al., 1989]、好酸好熱性細菌[Denda et al., 1988a, 1988b, 1989, 1990]より相次いで単離された F_0F_1 -ATPase 様酵素の F_1 相当部分は、2 つ大きな α 、 β という構成成分を持つ点では F_0F_1 との類似性を見ることができたが、 α が 60 k と F_1 の α (55 k) に比べ非常に大きいこと、 F_1 の特異的な阻害剤であるアザイドで阻害されないなどの決定的な違いを有していた。伝田らが好酸好熱性細菌、*Sulfolobus acidocaldarius* より単離した遺伝子の解析の結果、この ATPase の主要サブユニットは、 F_1 のそれと 25% 程度の相同性しか持たない一方で、ほぼ同時期に塩基配列が決定された真核生物の空胞系に発現している V (vacuole) タイプの ATPase と 50% 程度の高い相同性を持つことが分かった[Gorgarten et al., 1989]。このことは、古細菌には F (F_0F_1) タイプではなく、V タイプの ATPase のみが存在するという説を産むきっかけを作る。

1.3 古細菌と真正細菌の分化と V/F タイプ ATPase の分化との関係

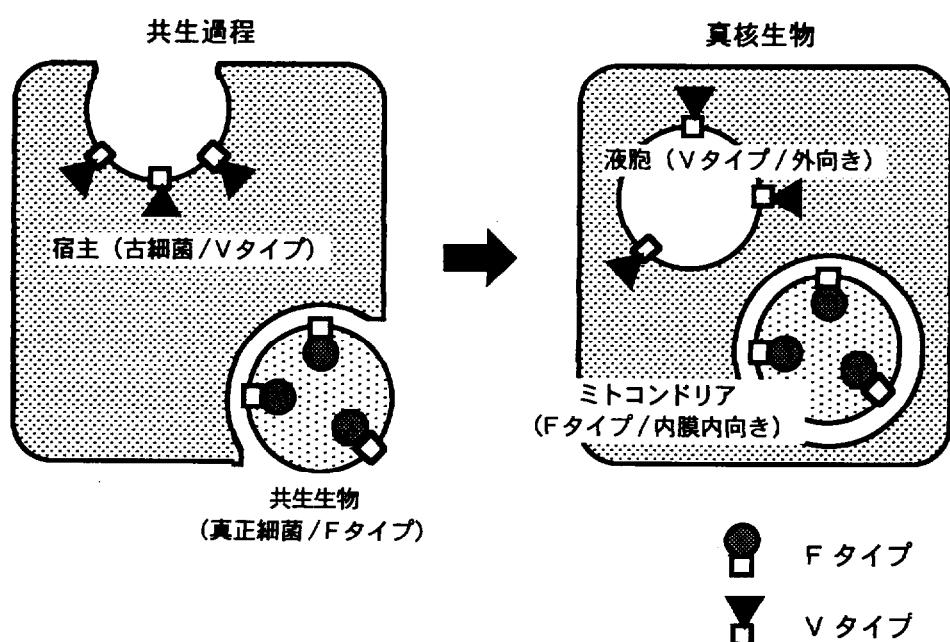
真正細菌には F タイプのみ、古細菌には V タイプのみが分布すると仮定すると、F、V 両タイプ ATPase に共通の始原タイプ ATPase を持つ始原生物が真正細菌と古細菌に分化、その後の進化を経て

真正細菌の ATPase は現在の F タイプに、古細菌の ATPase は現在の V タイプに変化したという仮説が成り立つ[Gogarten et al., 1989] (図02)。



(図02) 古細菌と真正細菌の分化と F / V タイプの分化時期が同じと考えたときの進化の仮説を模式か下もの。始原タイプの ATPase を継承した古細菌は進化の途中でそれを V タイプに変化させ、真正精勤は F タイプに変化させたと考えられる。

この説は、Margulis らの細胞内共生による真核生物発生説[Margulis, 1981]をうまく説明するため適正なものと思われた。つまり、共生の際、宿主となる古細菌に、ミトコンドリアや葉緑体になる真正細菌が共生したと仮定すると、宿主のプラズマ膜の陥入により派生したと思われる空胞系の V タイプ、共生生物のプラズマ膜由来と思われるミトコンドリアなどの内膜に存在する F タイプの ATPase という具合に、分布もトポロジーも非常によく合致する (図03) [Nelson et al., 1989]。

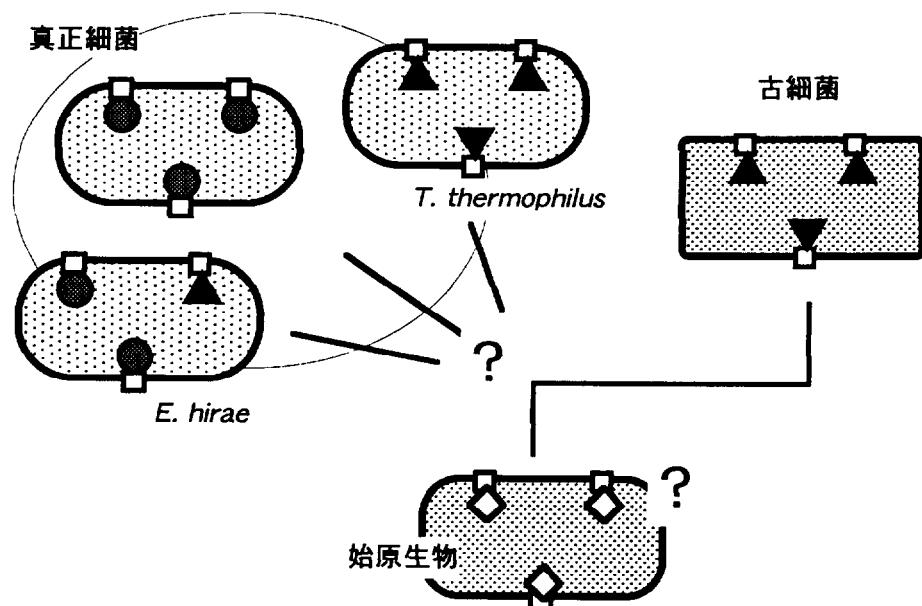


(図03) 宿主の古細菌は V タイプ、共生生物の真正細菌は F タイプのみを保持するとし、真核生物が細胞内共生によって生

じたときの過程を模式化した図。空胞系は宿主のプラズマ膜が貫入して進化したもの、ミトコンドリアや葉緑体は共生生物がエンドサイトーシスなどで細胞内に取り込まれたものとすると、現在の真核生物のVタイプおよびFタイプのトポロジーをよく説明できる。

1.4 真正細菌におけるVタイプの発見と真正細菌におけるF/V両タイプの共存

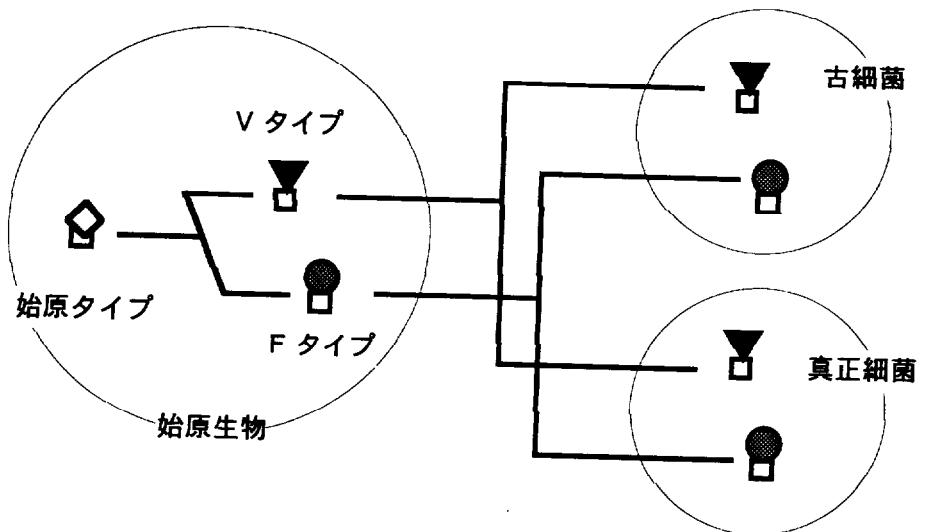
しかし、1990年、横山らにより、真正細菌に属する好熱菌、*Thermus thermophilus*より、明らかにVタイプに属すると思われるATPaseが単離される[Tsutsumi et al., 1991; Yokoyama et al., 1990, 1994]。この結果、前述のF/Vタイプの分化と真正細菌/古細菌の分化時期を同一とする仮説は修正を余儀なくされた（図04）。その後、柿沼らにより、すでにFタイプの存在が確認されている腸内球菌、*Enterococcus hirae*にもVタイプが存在することが報告された[Takase et al., 1993, 1994; Kakinuma & Igarashi, 1994]。ここに至り、少なくともVタイプは古細菌に限らず真正細菌にも分布しうること、そして、同一バクテリア内に両タイプが共存することも可能であることが示唆された。



〈図04〉 真正細菌にVタイプが見いだされたことにより不明確になった系統樹。真正細菌のVタイプが偶発的に古細菌のそれと同じようになったとは考えにくい。始原タイプの存在を仮定する必然性もなくなってしまう。

1.5 本研究の目的

*T. thermophilus*におけるVタイプの発見をうけ、我々はF/Vタイプの分布に対する認識を改める必要を感じ、またATPase進化仮説の修正を行なうべく準備をすすめた。具体的には、始原タイプのATPaseがVおよびFタイプに分岐したのは、始原生物が古細菌と真正細菌に分岐するはるか以前で、すでに始原生物中にて分化は完了していたこと。その後、両タイプのATPaseは、真正細菌および古細菌に継承され、あるものは両タイプを維持し、あるものは片方を欠落させたという仮説である（図05）。



〈図05〉新しく考えられた仮説。始原生物内ですでに始原タイプは遺伝子重複などにより、FおよびVタイプに分化、その後、真正細菌/古細菌への分化を経て現在に至るとするもの。これらな、真正細菌におけるVタイプの存在も矛盾無く説明できる。

本研究はこの仮説に則り、1) まだF/Vの報告のないバクテリアより、いずれか、もしくは両方のタイプのATPaseの存在を確認すること、2) すでにF/VいずれかのATPaseが報告されているバクテリアについても、もう一方のタイプのATPaseが存在し得ないかを探索すること、を通じてより多くのATPaseのタイプ別分布に関する情報を集めることを当初の目的とした。また、F/Vの広範な分布を示すために不可欠であるがまだ報告されていない、古細菌のFタイプATPase遺伝子の発見を最終的な目標とした。その結果、メタン産生菌よりFタイプの遺伝子を発見することができたのでここに報告する [Sumi et al., 1997]。

第2章 PCR 法によるバクテリアの F/V タイプ ATPase 遺伝子の確認

1990 年初頭、より安価なサーマルサイクラーの普及により、PCR 法を用いた遺伝子の単離が一般の研究室単位でも行なえるようになった。この系を応用し、F / V タイプの遺伝子存在の有無を、DNA 断片の増幅の有無を手がかりに手軽に判定できる系を構築した。その結果、いくつかの成果があがり有効性を示す一方で、PCR 法一般にいわれるのと同じ問題もはらんでおり、この系の運用には注意が必要であることも判明した。

2.1 方法

2.1.1 バクテリア

バクテリアは、生菌、集菌された凍結菌体、調製されたゲノム DNA のいずれかの形態で恵与されたものを使用した（表01）。*E. coli*、*P. denitrificans*、*T. aquaticus* は恵与された生菌を培養し、集菌後、ゲノム DNA を調製した。*P. modestum*、*M. barkeri* は凍結菌体よりゲノム DNA を調製した。

strain	gifted form	gifted from
<i>Escherichia coli</i> JM109	cell	
<i>Salmonella typhimurium</i> LT2	DNA	f
<i>Bacillus subtilis</i> NB22	DNA	c
<i>Staphylococcus aureus</i> 209P	DNA	f
<i>Paracoccus denitrificans</i>	cell	c
<i>Nitrosomonas europaea</i>	DNA	c
<i>Vibrio parahaeckelii</i> AQ3334	DNA	f
<i>Thiobacillus ferrooxidans</i> Fe1	DNA	c
<i>Propionigenium modestum</i>	DNA	b
<i>Enterococcus hirae</i> ATCC9790	DNA	g
<i>Thermus thermophilus</i> HB8	DNA	a
<i>Thermus thermophilus</i> HB27	DNA	a
<i>Thermus aquaticus</i>	cell	a
<i>Sulfolobus acidocaldarius</i> strain 7	DNA	
<i>Halobacterium salinarium</i> L-33	DNA	c
<i>Methanosaeca barkeri</i> DSM800	cell	d

a, 本学旧大島・有坂研究室; b, Lehrstuhl für Mikrobiologie, Technische Universität München, München, F.R.G.; c, 本学旧山中・副森研究室; d, 産業医科大学学生化学教室; e, 本学正田研究室; f, 岡山大学薬学部微生物薬品化学教室; g, 千葉大学薬学部臨床化学研究室。

(表01)

2.1.2 培養

E. coli と *P. denitrificans* は、1.5% yeast extract、1.5% polypeptone、0.2% NaCl を含む培地を用い、37°C の好気的な条件下で培養した。*T. aquaticus* は、0.3% yeast extract、0.3% polypeptone、0.2% NaCl を含む培地で 80°C で培養した。得られた菌体は、それぞれただちに 2,000×g、4°C

で集菌し、ゲノムDNAの調製に用いた。

2.1.3 ゲノムDNA調製

E. coli、*P. denitrificans*、*T. aquaticus*、*M. barkeri*の菌体温重量10gを25mLのTEバッファ(10mM Tris-Cl pH 8.0、0.1mM EDTA)によく懸濁し、10% SDSを2.5mL加えて菌体を溶解させた。溶解を確認した後、等量のTE飽和フェノールを加え、30分間程やかに混和。6,000×g 10min遠心したのち、上清水溶液部分を別の容器に移し、再び等量のTE飽和フェノールを加え、15分程度穏やかに混和。遠心分離までの同様操作を3度繰り返し、最後に得られた水溶液に無水エタノールを3倍量重層し穏やかに混和した際生じるゲノムDNA沈殿をすくい上げ、70%エタノールで2度、無水エタノールで2度洗滌後、適量のTEバッファに溶解した。

2.1.4 PCR用プライマ合成

PCR法で用いるオリゴヌクレオチド(プライマ)は、F/VそれぞれのタイプのATPase触媒サブユニット(Vタイプは α 、Fタイプは β)のアミノ酸配列から非常に高い保存性を保つ領域を3カ所選びだし、これらをもとに設計した(VタイプはV1、V2、V3、FタイプはF1、F2、F3とそれぞれ名付ける;図06)。鎖長は30塩基弱と、通常PCRで使われるものより比較的長くし、特に伸長反応方向に位置する3'端の配列についてはすべてのコドンの組み合わせに対応できるよう混合物として作成するなど、目標となる塩基配列が存在したときに取りこぼしのないよう配慮した(図07a、b)。

V type		F type	
	== (V1) ==>		== (F1) ==>
Carrot 69	99-GNIEFDGIQKPLKT-111	Bovine mitochondrion	154-GLFGGAGVGKTVF-166
<i>Thermus</i> α	81-NCIYDGIQQRPLER-93	<i>E. coli</i> β	147-GLFGGAGVGKTVN-159
<i>Sulfolobus</i> α	82-GKIYDGLQQRPLDS-94	<i>Rps. blastica</i> β	150-GLFGGAGVGKTVL-162
<i>E. coli</i> β	87-GRMNVLGEVDM-99	<i>Sulfolobus</i> α	220-AIPPFPSST-232
	<== (V2) ==		== (F2) ==>
Carrot 69	274-VVYVGCGERGNEM-286	Bovine mitochondrion	186-GVGERTREGNDLY-198
<i>Thermus</i> α	250-VVYVGCGERGNEM-262	<i>E. coli</i> β	179-GVGERTREGNDFY-191
<i>Sulfolobus</i> α	255-VIYVGCGERGNEM-267	<i>Rps. blastica</i> β	182-GVGERTREGNDLY-194
<i>E. coli</i> β	172-SIFAVVTRIG-187	<i>Sulfolobus</i> α	259-SCGNMTDEL-271
	<== (V3) ==		<== (F3) ==
Carrot 69	318-TSNMPVAAREASI-330	Bovine mitochondrion	308-IYVPADDLTDPAP-320
<i>Thermus</i> α	291-TSNMPVAAREASI-303	<i>E. coli</i> β	295-VYVPADDLTDPAP-309
<i>Sulfolobus</i> α	296-TSNMPVAARESSI-308	<i>Rps. blastica</i> β	304-IYVPADDLTDPAP-316
<i>E. coli</i> β	208-QMELPGNLRVVA-220	<i>Sulfolobus</i> α	389-VSPGGFVT-401

〈図06〉 V1~3、F1~3 設計のもととなるアミノ酸配列

V type

```

V1: -(left)=>
    IleTyrAspGlyIleGlnArgProLeu
  5'-ATTTATGATGGGATTCAACGTCTTT -3'
      G A C   CC
      A   A
      G   G

V2: <-(right)=
    TyrValGlyCysGlyGluArgGlyAsn
  5'-TATGTTGGTTGGAGAACGAGGAAA -3'
      C   C   C   C
      A   A
      G   G

  3'-ATTTATGATGGGATTCAACGTCTTT -5'
      G   G   G   G
      T   T
      C   C

V3: <-(right)=
    AsnMetProValAlaAlaArgGluAla
  5'-AATATGCCCTGTTGCAGCACGAGAACG -3'
      C   C   C
      A   A
      G   G

  3'-TTATACGGACAACGTCGTGCTCTTCG -5'
      G   G   G
      T   T
      C   C

```

図07a) V1、V2、V3 の塩基配列

F type

```

F1: -(left)=>
    GlyGlyAlaGlyValGlyLysThr
  5'-CGTGGTGCTGGTGTTGGTAAAC -3'
      C   C   C   C   C   C   G
      A   A   A   A   A   A
      G   G   G   G   G   G

F2: -(left)=>
    GluArgThrArgGluGlyAsnAsp
  5'-GAACGTACTCGTGAAGGTAAATGA -3'
      G A   C   C   C   G   C   C
      A   A   A   A
      G   G   G   G

F3: <-(right)=
    TyrValProAlaAspAspLeuThrAsp
  5'-TATGTTCCCTGCTGATGATTAACTGA -3'
      C   C   C   C   C   C   C   C
      A   A   A   A   G   A
      G   G   G   G

  3'-ATACAAGGACCGACTACTAGATTGACT -5'
      G   G   G   G   G   G   G   G
      T   T   T   C   T
      C   C   C   C

```

図07b) F1、F2、F3 の塩基配列

反応に用いるのは V1-V3、V1-V2、F1-F3、F2-F3 の組み合わせで、もし対応するタイプの遺伝子が存在すれば、それぞれ、669 bp、543 bp (S. acidocaldarius の場合)、492 bp、384 bp (E. coli の場合) 程度の長さをもつ DNA 断片の増幅が期待できる。

2.1.5 PCR

反応には、主に Taq ポリメラーゼ (BIOTECH) を用いた。反応液の総量は 100 μL で、主要な成分は、0.2 μg 程度のゲノム DNA (テンプレート)、各 100 pmol のプライマ (V タイプ検査には V1-V3、F タイプ検査には F1-F3)、10 μL 10×反応溶液 (キット付属)、1.5 mM MgCl₂、各 1.25 mM デオキシヌクレオチド (dATP、dCTP、dGTP、dTTP)、2.5 単位の Taq ポリメラーゼである。混合反応液には、反応途中での水分蒸発を防ぐためにミネラルオイルを適量滴下した。増幅反応はアステッ

ク社製サーマルサイクラ（PC-700）を用いて行なった。温度調節は1サイクルにつき、94℃、55℃、74℃をおののおの1分ずつ、合計30サイクルのプログラムを利用した。增幅の確認は1%アガロースゲル電気泳動の後、エチジウムプロマイド染色を行ない紫外線下で写真撮影した。

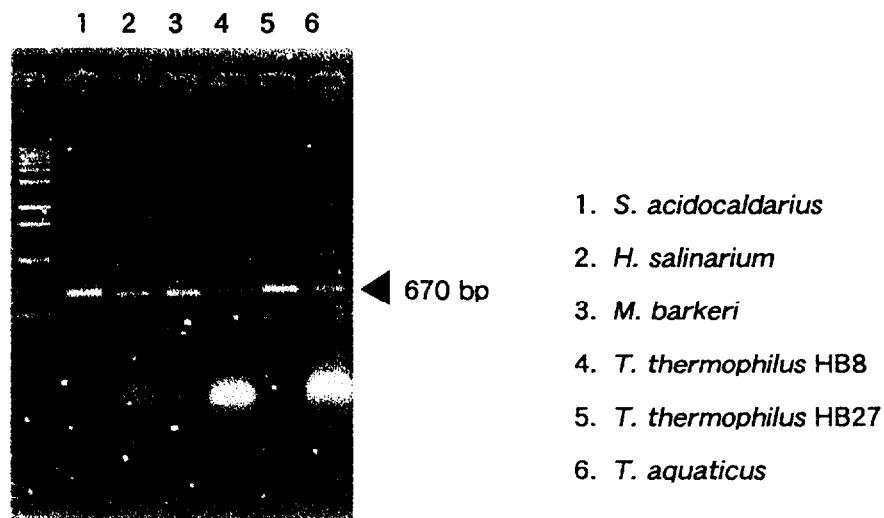
2.1.6 確認のためのPCR

2回目の確認のための増幅に備え、1回目の増幅産物はゲルから切り出し精製の後、TEバッファに溶解させた。PCRは、テンプレートとして切り出し精製した増幅産物（約数ng）、プライマとしてV1-V2もしくはF2-F3を用いることを除いて、1回目の増幅と同じである。

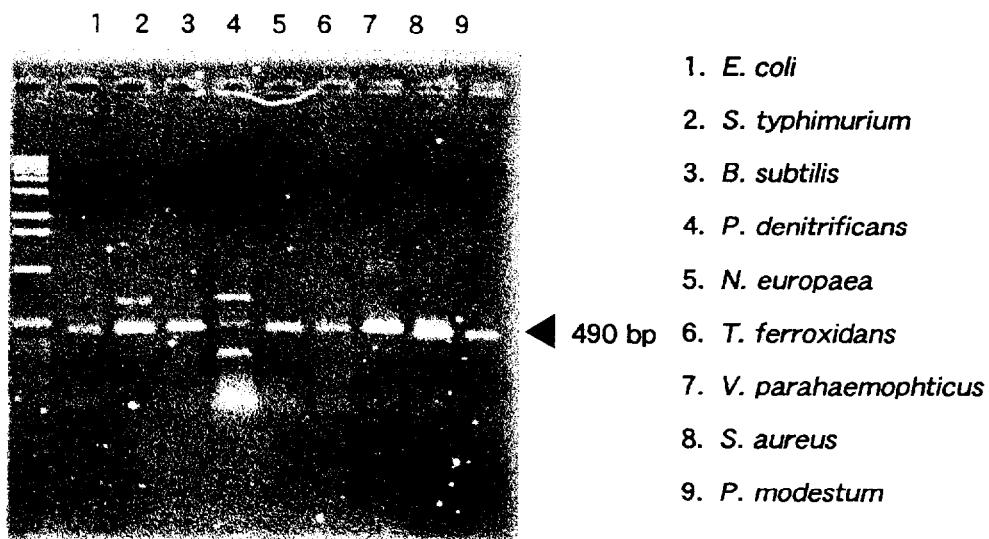
2.2 結果

2.2.1 既知のVタイプ、Fタイプ遺伝子の確認

従来の報告よりVタイプが存在することが分かっている古細菌および*Thermus*属についてVタイプのテストを、Fタイプが存在することが予想される、もしくは知られている*Thermus*属以外の真正細菌についてはFタイプのテストを、それぞれポジティブコントロール実験として行なったところ良好な結果が得られた（図8、9）。



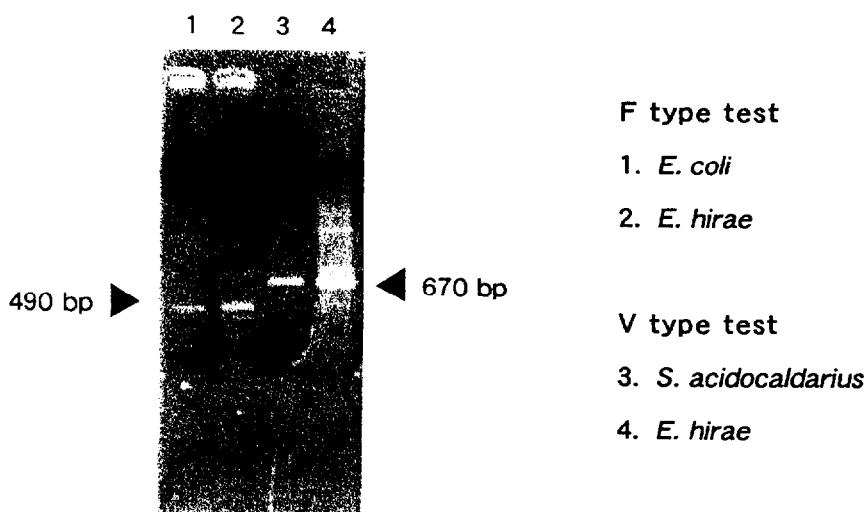
〈図08〉 Vタイプテストのポジティブ実験。Vタイプ遺伝子が存在すれば、黒三角で示した670 bp付近にバンドが生じる。



(図09) Fタイプテストのポジティブ実験。Fタイプ遺伝子が存在すれば、黒三角で示した 490 bp 付近にバンドが生じる。

2.2.2 *E. hirae*におけるV/Fタイプ遺伝子の確認

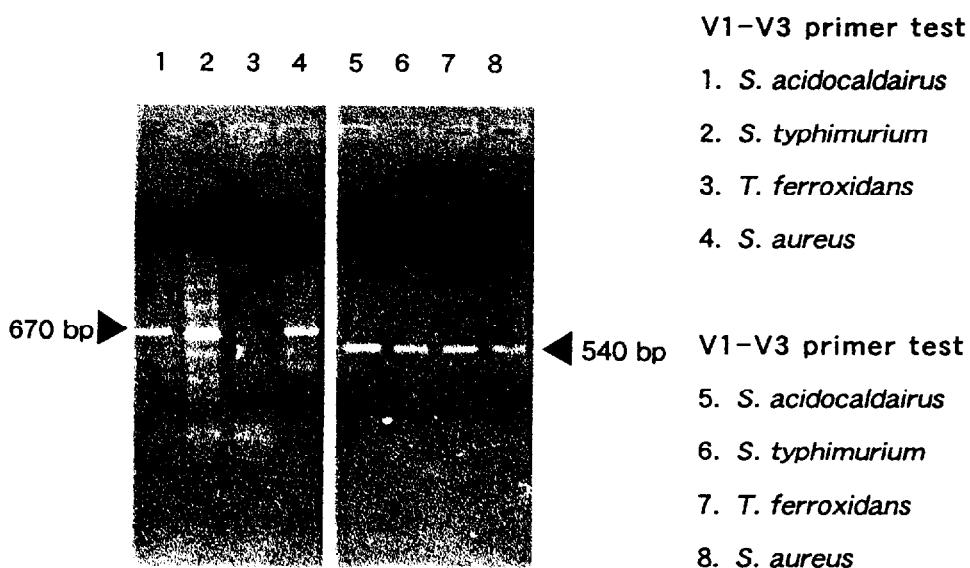
E. hirae は、呼吸鎖を持たず膜電位の維持には通常FタイプのATP水解反応を用いている。しかしこの酵素の至適pHは酸性であること、あるタンパク質を発現できない変異株はアルカリ性条件下では生育できること、などから、アルカリ条件下では別のナトリウムイオン輸送性ATPaseを発現して $\Delta \mu \text{Na}^+$ により膜電位を維持している可能性が柿沼らにより示唆された[Kakinuma et al., 1990]。このナトリウムイオン輸送性ATPaseは精製標品のサブユニット組成や阻害剤に対する挙動（アシドで阻害されず、硝酸イオンで阻害される）などからVタイプであると予想されたが当初は遺伝子による確認はされていなかった。柿沼らにより恵与された*E. hirae*のゲノムDNAを用いてFおよびVタイプのテストを行なったところ、両タイプにおいて増幅が見られた(図10)。



(図10) すでにタンパク質レベルで存在が確認されている*E. hirae*における、F/Vタイプのテスト。Vタイプのポジティブ標品には*S. acidocaldarius*、Fタイプには*E. coli*を用いた。Fタイプ、Vタイプ遺伝子の存在を示す、それぞれ490、670 bp のバンドが確認できる。

2.2.3 *Thermus* 属以外の真正細菌におけるVタイプ遺伝子の確認

すでにVタイプの報告されている *T. thermophilus* HB8 とその同属バクテリアを除いた真正細菌においてVタイプ遺伝子が存在する可能性がないか検査を行なった。その結果、*S. aureus*、*S. typhimurium*、*T. ferrooxidans*において 690 bp 付近の増幅が確認された（図11、レーン2～4）。それぞれの増幅 DNA 断片を切り出し精製し、テンプレートに用いた2回目のPCR（V1-V2 プライマを使用）でも 540bp 付近に増幅が確認された（図11、レーン6～8）。



〈図11〉 真正細菌のVタイプテスト。コントロール標品には *S. acidocaldarius* を使用した（レーン1、5）。1回目の増幅では非特異的バンドがあつて不明瞭でも、2回目の増幅（レーン6～8）では単一のバンドで確認できる。

2.2.4 古細菌および *Thermus* 属におけるFタイプ遺伝子の確認

古細菌および *Thermus* 属ではVタイプの存在しか示されていない。まず、真正細菌である *Thermus* 属においては、*E. hirae* のようにV/F両方の遺伝子を保持する可能性が高い。このような予想通り、*Thermus* 属からはFタイプ遺伝子由来と思われる増幅が確認できた。さらに古細菌からもメタン產生菌、*M. barkeri* と好塩菌、*H. salinarium* より F由来の増幅が確認できた。いずれの増幅も切り出し精製し、それをテンプレートとして用いた2回目の増幅（F2-F3 プライマ使用）においても 380 bp 付近の単一のバンドが確認された（図12）。

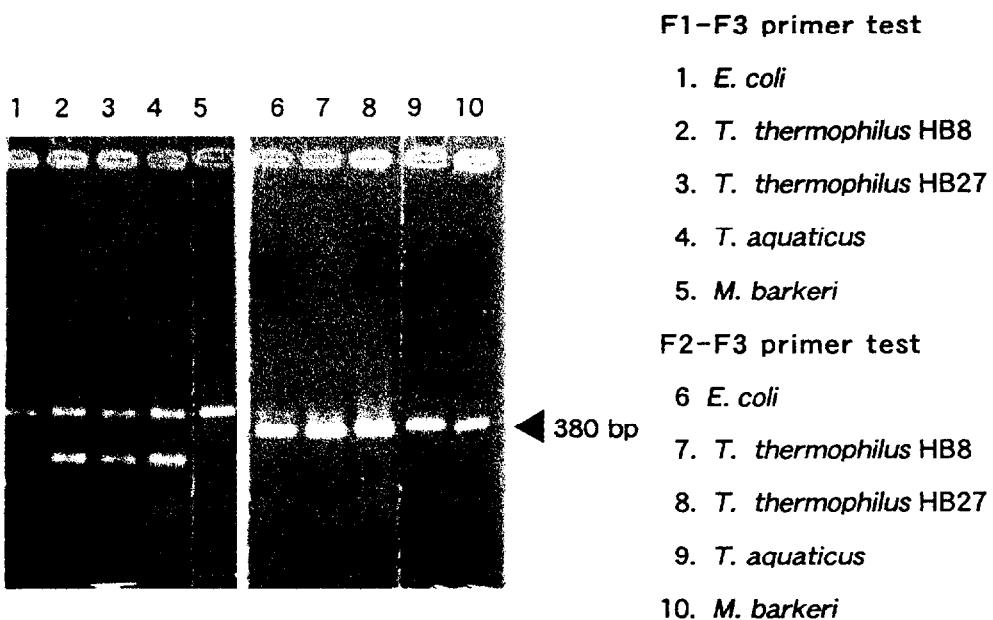


図12) *Thermus* および古細菌におけるFタイプテスト。*H. salinarium* の結果は示していない。

2.2.5 増幅産物の塩基配列確認

新規に確認された真正細菌のVタイプのうち、*S. aureus* のものをpUC119に組み込んでサブクローン化し、その塩基配列を決定したところ、増幅が期待される領域のアミノ酸配列と相同性を持つコドンの並びを確認できた。新規に確認された古細菌のFタイプ、*M. barkeri* および *H. salinarium* より得られた増幅断片も同様に塩基配列決定を行なったところ、*M. barkeri* ではやはり期待されるアミノ酸配列と相同性を持つコドンの並びが確認されたが、*H. salinarium* では、偶然、同じ長さで2回目のPCRにおいても同じ挙動をする、まったく関係のないゲノムDNA断片が増幅されていることが判明した。

2.3 考察

2.3.1 検出系の長所と欠点、考え方の改善策

PCRで未知のDNA断片を増幅し単離する場合、増幅したいDNA断片の両端の塩基配列、20 bp程度が分かればよい。ある遺伝子が、たがいにそれほど離れていない（理想的には1~2 kbp程度の距離を持つ）2カ所以上の比較的高い保存性領域保持する場合、そこから予想されるDNA配列よりプライマを設計しPCRを行なうことで、その遺伝子の一部を単離することができる。逆に、対象となる遺伝子の保存領域において高度な保存性が期待される場合、PCRによるDNA断片増幅の有無により、その遺伝子の存在の有無を確認することができるはずである。FおよびVタイプATPaseはアミノ酸レベルで高い相同性を保持する領域をそれぞれ数カ所ずつ保持するため、この系によりその遺伝子の存在の有無を確認することが容易であると考えた。

間接的に酵素、もしくはその遺伝子の存在を確認するには、酵素独自の活性を測定する方法、プローブに抗体を用いるウエスタンプロット解析、プローブにDNA断片を用いるノザン/サザンプロット解

析などが考えられるが、いずれも状態のよい、比較的まとまった量の菌体や核酸などを必要とする上、標品調製および検出作業が煩雑で、多種に及ぶ標品の検定には向かない。それに対し、PCR法による検出システムでは、微量のゲノム DNA を手に入れることができれば、あとは必要な試薬を混合し反応を行なうだけで、簡便な検出が可能な点でアドバンテージを持つ。

一方で、PCR では、コンタミネーションやプライマの非特異的吸着によるアーチファクトの可能性に加え、テンプレートに用いるゲノム DNA の GC 含有量やそれに関連する高次構造による増幅障害などの可能性を考慮する必要がある。コンタミネーションについては、ゲノム DNA 標品への混入を除けば、反応系のネガティブコントロール実験（テンプレート DNA を加えずに反応を行なう）を定期的に行なうことによって誤った判断をくだす危険性を排除できる。ゲノム DNA 標品への異物混入の検出は難しいが、小分けするなどして入手時のクオリティを保持する配慮は必要である。

V/F タイプの触媒サブユニットには、25% 程度はあるが相互に類似性がある。F タイプの検出プライマで V タイプを、逆に V タイプの検出プライマで F タイプを増幅してしまわないようにプライマの設計においては可能な限り共通の保存領域は避けた。また、V にはあって F にはない挿入配列を挟んでプライマの設計することで、誤った増幅が行なわれても、その長さで判断が付くように工夫した。他方、非特異的増幅による判断ミスを避けるため、第 3 のプライマを用意し、1 回目の増幅断片をテンプレートに 2 回目の増幅を確認し、都合 3 力所の類似性を確認するシステムにした。

このチェック機構はおおかたうまく機能したが、残念ながら *H. salinarium* での F タイプの増幅をサブクローニングし塩基配列を決定したところ、まったく関係のない領域が偶然、同じ長さで増幅していたことが分かった。この断片は、内部に第 3 のプライマとの相同領域も持つため、その塩基配列を決定するまで正常な増幅と区別することは難しい。このようなアーチファクトを産んだ背景には、高い GC 含有のテンプレートと混合プライマを用いた際の PCR の難しさがある。

一般に GC 含有量の高い DNA 断片の PCR 増幅は完全一致のプライマを用いても、その設計によつては増幅がうまくゆかず、対処に苦慮することがある。近年、超好熱菌より単離されたポリメラーゼを使用した様々なアプリケーションが提供されており、以前よりは状況はよくなつたが、いずれも一般的な反応系において増幅が見られない際に試してみる価値はあるものの、決定打にはなり得ていない。従って、本検出系においても、増幅が無いことがすなわち遺伝子の存在を否定する結論には結びつかない。簡便な検出系という主旨からは外れるが、遺伝子の存在を否定する決定的な手段は、最近積極的に進められている全ゲノム塩基配列決定のみに限られる。

実際、本検出系においても GC 含量の問題は既知の遺伝子の増幅の際もたいへんな問題となった。GC 含量が非常に高い（65% 以上、コドン最終塩基では 90% 近く）*H. salinarium*、*Thermus* 属においては安定な増幅を得るのはいかなる状況においても難しい。先に触れた超好熱菌由来のポリメラーゼ

を用いたアプリケーションをはじめ、2価カチオン量、プライマ量のスクリーニング、テンプレートの断片化、アルカリ変性処理などを用いた一本鎖化など多くの手法を試したが、安定な増幅は期待できなかつた。唯一、特異的な増幅が見られたのは、サーマルサイクル不調で温度調節がプログラム通りにゆかなかつたときであるが、残念ながらこれを正常なマシンで再現することはできなかつた。

他方、やはり GC 含量の高い好塩菌の一種、*Haloferax volcanii* を用い、報告されている V タイプの塩基配列[Steinert et al., 1995]より、本検出系で設計した V タイプの検出用プライマとまったく同じ場所のプライマを作製し PCR 反応を行なつたところ、安定な増幅を得ることができた。本検出系に用いた混合プライマの中には、このとき用いた完全一致のプライマと同じ配列のものも存在しているはずだが、安定した増幅は見られない。*H. volcanii* において増幅に十分な量の完全一致のプライマと、本検出系のプライマを混合して増幅したところ、混合比率によっては増幅が見られなくなることが分かった。これは反応条件によっては、多く存在する増幅には使用されないプライマが、極わずかに存在する適正なプライマによる増幅反応を阻害している可能性を示唆する。

本検出系は現在のままでも、中～低度の GC 含量のゲノム DNA を持つバクテリアには有効に機能する。高度に GC に富んだ配列を有する標品用に改良するとすれば、コドン第 3 塩基を GC に限った混合プライマを作製する、より適正な温度コントロールプログラムを考案するなどの工夫が必要であろう。

2.3.2 新奇 V/F タイプ遺伝子存在の可能性

前述の通り、問題は多く残されているが、現時点では本検出系により真正細菌で 3 種の V タイプ、古細菌および *Thermus* 属でつごう 4 種の新奇 F タイプ遺伝子の存在が示唆された。本研究で取り扱つた真正細菌の数は現存しうるバクテリアの種類からすれば微々たるものであるにも関わらず、従来報告されている *T. thermophilus*、*E. hirae* 以外にも、塩基配列確認の済んだ *S. aureus*、そして他の 2 種の真正細菌に V タイプが存在する可能性が示唆されたことは、V タイプの真正細菌における分布を裏付ける上でも重要な情報を提供する。

また、古細菌の *M. barkeri* において F タイプ遺伝子の存在が確認されたことは、たいへんな成果である。後に分かったことだが、*M. barkeri* と大変近縁の *M. mazei* でのサザンプロット解析の結果 [Müller, 私信] と、近年全ゲノム解析を終えた、*Methanococcus janaschii*、*Methanobacterium thermoautotrophicum*、*Archaeoglobus fulgidus* という 3 種の古細菌における類似性検索[<http://www.tigr.org/tdb/mdb/mdb.html>]の結果から、いずれも現時点では F タイプの存在を示唆する情報は得られていない。メタン产生菌として *M. barkeri* が入手可能であったこと、そのゲノム DNA の GC 含量が 40% 程度[Boone et al., 1993] と低かったため本系標品に向いていたことが幸いしたもの、古細菌において限られた種にしか存在しない F タイプの遺伝子を検出できたことは、本検出系の有効性を示すよい例となつた。

第3章 古細菌からのFタイプATPase遺伝子の単離と塩基配列決定

PCR法を用いた検出系により、古細菌に属するメタン産生菌、*M. barkeri* にFタイプ遺伝子が存在する可能性が示唆された。この増幅断片をプローブに古細菌のFタイプ遺伝子クラスタを単離し、その塩基配列を決定した。すでに、このメタン産生菌にはVタイプの遺伝子が存在することが報告されていたが、それとはまったく異なるFタイプの遺伝子であることが明らかにされた。この古細菌のFタイプ遺伝子には、 δ サブユニットをコードする遺伝子が存在しない、 b サブユニットが異常に大きい、 γ サブユニットがアミノ末端とカルボキシル末端とで別々の遺伝子にコードされている、などの特徴を持つ。残念ながらこの遺伝子の産物であるタンパク質や mRNA の検出は失敗に終わったため、本遺伝子が機能を持つかどうかは現時点では明らかでない。

3.1 方法

3.1.1 バクテリアとその培養

M. barkeri MS（基準株、DSM800）は、共同研究者の古賀らが Deutshce Sammlung für Mikroorganismen und Zellkulturen (DSM) より得、継代培養しているものを使用した。培養は絶対嫌気的条件下（気層を 80% 窒素+20% 二酸化炭素にて置換）で行ない、0.5% メタノールを含む DSM120 (DSM が推奨するメタン産生菌用培地；pH 6.85、ナトリウムイオン濃度 約 40mM) を培地に用了。pH の異なる培地はカリウムリン酸バッファ、ナトリウムイオン濃度の異なる培地は加える NaCl 量で調節し作製した。

3.1.2 プローブのラベルと遺伝子の単離、塩基配列決定

PCR 増幅産物はアガロースゲル電気泳動後、切り出し精製を行ない、クローニングベクター pBluescript II SK- (Stratagene 社) の EcoRV サイトに挿入し、サブクローン化した。このプラスミドを大腸菌 JM109 株に導入のち大量に得、EcoRI および HindIII により切断して切り出した DNA 断片をプローブとして使用した。プローブのラベルおよびサザン/コロニーハイブリダイゼーションは ECL ランダムプライマラベルシステム (Amarsham 社) を使用した。

*Bam*HI で *M. barkeri* ゲノム DNA を切断し、8 kbp 付近の断片をアガロースゲル電気泳動と切り出し精製法を用いて回収した後、pBluescript II SK- ベクターの *Bam*HI サイトに組み込んだものをコロニーハイブリダイゼーションのライプラリに使用した。ポジティブクローンを選び、培養してプラスミドを得たのち、*Bam*HI で切り出されるベクター以外の挿入断片を切り出し精製法で回収した。挿入断片は超音波により適当な長さに破碎した後、600 bp 付近の断片を平滑化、クローニングベクター pUC18 の *Sma*I サイトへのライゲーション作業を経て大腸菌に導入した。無作為に選んだコロニーからプラスミドを精製し、シークエンス反応のテンプレートとして用いた。

3.1.3 塩基配列の解析

シークエンス反応の結果は、随時 ATSQ ソフトウェア（自動塩基配列結合ソフト；Software Development 社）による処理を行ない、DNA 両方向からの配列決定を確認して作業を終えた。遺伝子の読み枠予測は GENETYX-Mac (Software Developement 社) の機能を用いた。既知の遺伝子との類似性検索には、および、アライメントの作成は Genetics Computer Group (GCG ; Wisconsin 大学) の提供するパッケージを使用した。

3.2 結果

3.2.1 遺伝子の単離

PCR で得られた 380 bp の DNA 断片（図13、黒い太線B上の白抜き部分）は、サザンプロット解析の結果、*M. barkeri* の *Bam*HII 切断産物の 8 kbp 付近で交叉した（図14）。約 100 クローンの 8 kbp 付近の *Bam*HII 切断産物で構成されるプラスミドライプラリーをスクリーニングしたところ、1 個のポジティブクローンを得ることができた（図13、黒い太線B）。塩基配列を決定したところ、このクローンには α のカルボキシル末端と γ 、そして δ サブユニットをコードする遺伝子を除き、F タイプ構成成分遺伝子のほとんどすべてが含まれていることが分かった。そこで、この *Bam*HII 断片の 3' 端に位置する約 800 bp の *Sac*I 断片（図13、E 上の白抜き）を切り出し精製し、新たなプローブとして染色体歩行を行なった。ハイブリダイゼーションの結果、*Eco*RI 切断産物で構成されるライプラリに残りの遺伝子をすべて網羅するのに十分な長さ（約 4 kbp）の交叉が確認されたので、4 kbp 程度の長さを持つ *Eco*RI ライプラリを作成、スクリーニングを行なった。約 400 クローン中より、4 つのポジティブクローンを単離した。これらの両端の塩基配列を決定したところ、同一かつ先に単離した *Bam*HII 断片と重複を持つ目指すクローンであると判断された（図13、黒い太線E）。このクローンについても *Bam*HII 断片と同様な方法でシークエンスを行なった。塩基配列決定の結果、残りの *Eco*RI 断片には予想通り、残りの F タイプ遺伝子および、転写終了領域と思われる配列を含んでいることが分かった。得られた両断片の塩基配列は、GenBank データバンクに AF028006 として登録した。

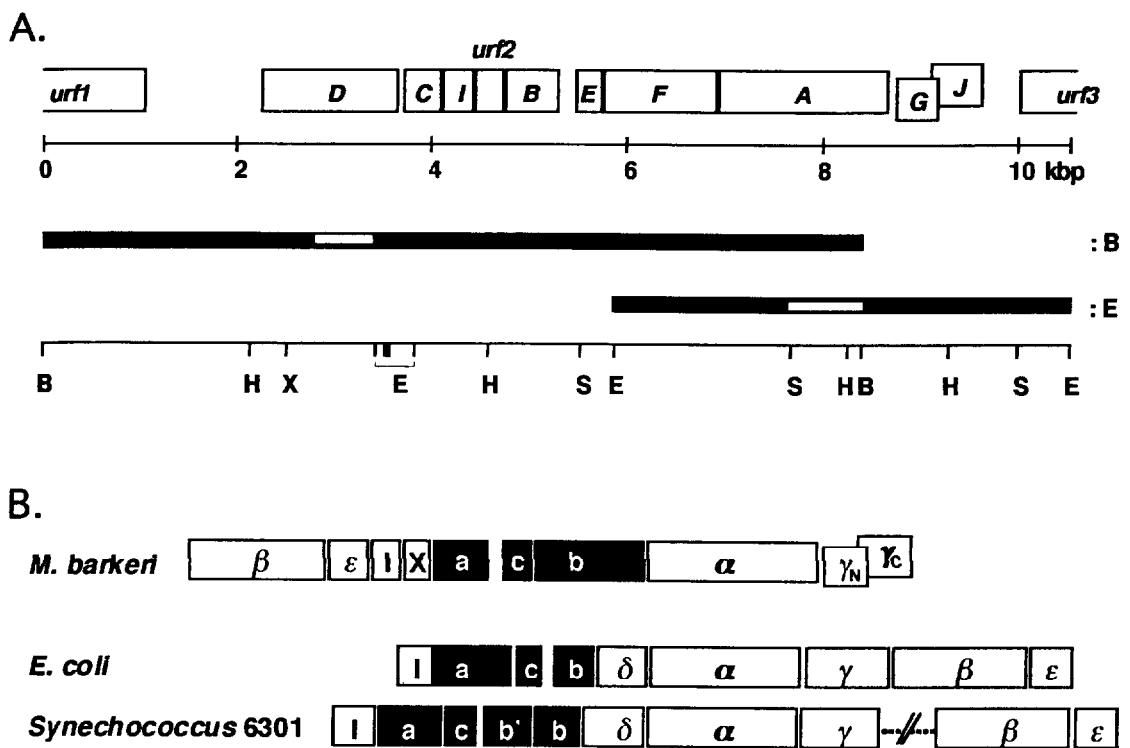


図13) *M. barkeri* の *atp* オペロンと制限酵素サイト地図。Δ. *urf* は機能不明の読み枠、白枠内のアルファベット A、B、C、D、F、G、I、J は遺伝子名を示す。太線 B、E はそれぞれ単離した 8 kbp *Bam*HI 断片、4 kbp *Eco*RI 断片。太線内の白抜きはそれぞれの断片を単離するのに使用したプローブの位置を示す。B, *Bam*HI; E, *Eco*RI; H, *Hind*III; S, *Sac*I; X, *Xba*I。B. *M. barkeri*, *E. coli*, *Synechococcus* 6301 の *atp* (*unc*) オペロン内の遺伝子の並び。枠内のアルファベットおよびギリシャ文字はサブユニット名称を表わす。I は 14k タンパク質、X は *urf2* 産物、 γ_N および γ_C はそれぞれ γ サブユニットのアミノ末端側、カルボキシル末端側との相同性を持つ遺伝子産物であることを示す。黒いボックスは F_o のサブユニット。

3.2.2 *M. barkeri* の F タイプ遺伝子の構成

単離した 2 つの DNA 断片の全塩基配列を決定した結果、その GC 含量は 43% と文献値[Boone et al. 1993]とほぼ一致した。これらのクローニングには、10 個の遺伝子と思われる読み枠が存在し、それぞれ順に、*atpD*、-C、-I、*urf2*、*atpB*、-E、-F、-A、-G、-J と名付けた。*urf* は未知の読み枠の意で、完全ではないが得られたクローニングの 5' 端と 3' 端にそれぞれ読み枠らしき配列（仮に *urf1*、*urf3*）を見いだせたため番号は 2 となっている。他の遺伝子の頭についている *atp* は初期に *unc* と名付けられそのまま定着した大腸菌を除く F タイプのオペロンの一般名称である。*atpJ* 以外の各サブユニットをコードすると予想される遺伝子の名前（アルファベット）は慣例に従った（表02）。読み枠予想は、各フレームにおける終止コドン (TAA, TAG, TGA) の出現位置を視覚化し、開始コドン (ATG) との組み合わせで行なった（図14）

Gene	Subunit	Amino acids	Molecular weight	
			<i>M. barkeri</i>	<i>E. coli</i>
<i>atpD</i>	β	469	51,891	50,286
<i>atpC</i>	ϵ	144	15,846	15,051
<i>atpI</i>	(14k)	112	12,801	14,055
<i>urf1</i>	(unknown)	104	11,903	—
<i>atpB</i>	a	190	21,140	30,285
<i>atpE</i>	c	91	9,406	8,264
<i>atpF</i>	b	386	44,560	17,230
<i>atpA</i>	α	588	65,203	55,282
<i>atpG</i>	γ_N	149	16,696	—
<i>atpJ</i>	γ_C	186	21,348	—
—	γ	—	—	31,559
—	δ	—	—	19,328

〈表02〉

Initiation Codons : ATG, GTG,
Termination Codons : TAA, TAG, TGA,



〈図14〉 読み枠の予測に用いた図。終止コドンが現われないと空欄になるので読み枠の可能性が高まる。終止コドンは長い縦棒、開始コドンは小さな縦棒で示される。

atpD の直前（上流）および *atpJ* の直後（下流）には予想される読み枠が存在しないこと（図14参照）と、*atpD* の上流にメタン産生菌やその他の古細菌で報告されている転写開始部位（真核生物のTATA boxに相当；box A、box B、box IIと呼ばれる）[Allmansberger et al., 1988]と類似性のある

塩基配列が見つかったこと（図15、A）、*atpJ* の下流に転写終結を促す機能を持つと考えられる [Boone et al., 1993; Müllar, 私信] 回文構造および繰り返し構造が見られること（図15、B）などから、この遺伝子クラスタは転写単位、すなわち *atp* オペロンを形成していると予想される。

A.

```

ATAAGAGAGACTTGTGAAATAAGAAAAACTTGTGAAATAAGAAAAACTTGTGAAATAAGG
AGTTCTTACAAAACAAGCTTAGAAAGAGGGCACTCCAAACCGAATATAATGCCAGTT
          box A                                box B
GGGGGAGCTGTTAAAATAGCAATATAATGCCAGCTGGGGGAGCTGTTAAAM V D
          box II                               RBS      M   V   D
                                                (AtpD...)

```

B.

```

TAGCTCTGCCGGTGAAAAGCAGCGTGAGTCGGCAATGGTCCAGACAAAAATGATGAAT →
S   S   R   R   * (AtpJend)
CCGAAAAATGAGTTTATTGAGAAGTTTGAGAGGCTGAAACGGGAAGAGATTGTGAA ←←←
ATAAGAGAGACTTGTGAAATAAGAAAAACTTGTGAAATAAGAAAAACTTGTGAAATAAGG →→→

```

（図15）転写単位であることを示す領域の塩基配列。A. AtpD の上流にある古細菌で共通のプロモータ配列（box A, B, II）。B. AtpJ の下流にある回文および繰り返し構造。

すべての読み枠はリボソーム結合部位と思われる領域に続く ATG を開始コドンにすると仮定した。リボソーム結合部位は、16S rRNA の 3' 端の相補鎖であることが知られているので、*M. barkeri* の 16S rRNA 遺伝子[GenBank AF028692]から推測した。すべての遺伝子において、予想される開始コドンから 7 ~ 14 塩基の距離にリボソーム結合領域を見いだすことができた（表03）。AtpC のアミノ末端には特殊な繰り返し配列がある。すべての繰り返し単位に開始コドン、および、リボソーム結合配列を見いだすことができるが、この特殊な構造の意味は分からぬ（図16、A）また *urf2*, *atpG* では TAG が、*atpA* では TAA が、その他の遺伝子では TGA が終止コドンとして使われていた。*atpG* と *atpJ* との間で 103 bp もの長い読み枠重複が見られた（図16、B）ほかは、各遺伝子は互いに短い非コード領域で区切られていた。ただし、*atpA* と *atpG*, *atpB* と *atpE* はそれぞれ、55 bp と 150 bp の比較的長い非コード領域によって区切られている。

Gene	Sequence around RBS and start codon
<i>atpD</i>	CCAGCTTGGGGAGCTGTTAAA <u>ATGGTAGATCGAAGTCG</u>
<i>atpC</i>	CATGAATT <u>CAGGACTCATGAATT</u> CAGGACTCATGAATT
<i>atpI</i>	GCCGATTAGC <u>GGAGTT</u> CAGA <u>ATGTCTGACAAACC</u> CATCA
<i>urf2</i>	AAGAGGAAC <u>AGGAGGATT</u> ACA <u>ATGAATGAAATATTGAAT</u>
<i>atpB</i>	AATTGACAA <u>AGGAGGCCAGTAATGC</u> GACTTAGTCCTGAT
<i>atpE</i>	AACAAC <u>GGAAAGCACTATGG</u> CTTGGACACCTATAT
<i>atpF</i>	GAAATTGAG <u>AGGAAAGTAGACTATG</u> CTGATTGATTGGTT
<i>atpA</i>	AGTCCTAAC <u>TGTGGCCAAGATGAGATTAATCCC</u> AAACC
<i>atpG</i>	CGGAGACAA <u>AGGAAAATCATGACCGAAACC</u> ACTCAGAG
<i>atpJ</i>	AGCTTGCC <u>AGGTAAAGCTCAGGTATGGG</u> CTGTAGGGGA

（表03）各 *atp* 遺伝子の開始部位とそれに対応するリボソーム結合部位。下線が開始コドン、太字が予想されるリボソーム結合部位。

A.

```

GACGAAGCAATGGCCAAAAAATCAAGCCGGAAAATCATGAATTCAGGACTCATGAATT
RBS      M N S G L M N S

CAGGACTCATGAATTCAGGACTCATGAATCTTACGATTCTCTGCCTTCCAGGTCTTG
G L M N S G L M N S G L T I L L P F Q
                                         (AtpC...)

```

B.

```

(... AtpG)                                     (AtpJ start)
TIATGCGGICAGGGAACTGAAAGCTTGCCAGGTAAAGCTCAGGTATGGCTGTAGGGGA
Y A V R E L K A L P G K A Q V W A V G E
M G C R G

GCTTGTATTACGCCTGGAAAGATGAAGGCTTGCCGCTTATTGGACTTATAATGTGCCA
I V I H A W K M K A C R I L D F I M C Q
A C Y S R L E D E G L P L I G L Y N V P

AATTCGGTTAACGGCAATTACCCGCTTATAGCGCAAATTCTTGTGGAAAATGAAAAATT
I R I R Q L P R L * (AtpGend)
N S V K A I T P L I A Q I L V E N E K L
                                         (AtpJ...)

```

（図16）*M. barkeri* *atp* オペロンの特殊な構造。A. AtpC のアミノ末端は繰り返し構造を持つ。繰り返し単位を下線で、繰り返し単位中のリボソーム結合部位および開始コドンは太字で示した。B. ヤサブユニットをコードすると思われる遺伝子のオーバーラップ。AtpG 終了手前 110 bp から次の AtpJ のコードが始まっている。下線がオーバーラップ領域。太字は AtpJ 開始コドン。

3.2.3 他のFタイプ遺伝子との類似性比較

各遺伝子の予想されるアミノ酸配列は、*urf2*を除き、すでに報告されているFタイプのサブユニットのアミノ酸配列と何らかの類似性を持っていた。類似性検索の結果、AtpD (*atpD*から予想される翻訳産物；469aa)、AtpC (144aa)、AtpB (190aa)、AtpE (91aa)、AtpA (588aa) はそれぞれ、す

でに報告済みの F タイプの β (459aa; *E. coli* の場合。以下同様)、 ε (138aa)、 a (271aa)、 c (79aa)、 α (513aa) サブユニットと類似性を持っていた[GenBank M13704, LA2023, MTU64318, X99078]。AtpF (386aa) は、その前半の配列 (約120aa) で b サブユニット (156aa) との類似性を持つ[GenBank X99078]が、後半には他のどの遺伝子産物とも類似性を持たない余計な長いペプチド鎖が付加されていた。

ほとんどすべてのバクテリアの *atp* オペロンには、酵素の構成成分にならない約 14 kDa のタンパク質をコードする遺伝子がある。このタンパク質の機能はまだ分かっていない。*M. barkeri* の *atp* オペロンにも AtpI (112aa) をコードする遺伝子が見つかった。当初、類似性検索の結果では、AtpI が何に類似性を持つのかが分からなかったが、既知の 14 kDa タンパク質に AtpI を加えてアライメントしたところ、*Rhodospirillum rubrum* の 14 kDa タンパク質[SWISS-PROT, P15011]と 92aa でオーバーラップしそのうち 24% 程度の相同性を持つことが分かった。

一見すると、*M. barkeri* のオペロン内の遺伝子の並びは、*E. coli* [Downie et al, 1979, 1981]などがそうで、一般に典型的と考えられている遺伝子の並びである I - a - c - b - δ - α - γ - β - ε とはかけ離れているものに見えるかも知れない。しかし、未知および F_o の構成成分である a 、 c 、 b サブユニットをコードする遺伝子クラスタ、 α - γ のクラスタ、 β - ε のクラスタに分けて考えると、 δ がないこと、 γ が 2 つの遺伝子に分離している点などを除けば、 β - ε クラスタがオペロンの先頭に移動しているだけで、さほど特異的な配列ではなくなる。 β - ε クラスタがオペロンから分離し別の遺伝子座に位置したり、別の転写単位を形成することは、*Synechococcus* 6301 の *atp* 遺伝子などすでに報告されている（図13、B）[Cozens and Walker, 1987]。

3.2.4 主要サブユニットの保存性における特徴

F タイプにおいて、 β サブユニットの一次構造が種を越えて高度に保存されている点（60% 以上が同一アミノ酸残基）はその特性のひとつと言っても過言ではない。ところが、*M. barkeri* の β サブユニットはどの種の β サブユニット対しても 50% 程度と比較的低い相同性しか持たない（図17）。 α についても同様のことが言える。しかし、酵素活性に重要なヌクレオチド結合モチーフ（別名 Walker 配列；モチーフ A, GXGXXGKT およびモチーフ B, ZZZZD ; X は任意の、Z は疎水性のアミノ酸残基）[Higgins et al., 1986; Walker et al., 1982; Fry et al., 1986] は α 、 β とともに保存されていたり、 β の活性中心と見られる残基（ウシ心筋ミトコンドリア F タイプ β では 188 番目のグルタミン酸）が保存されるなど、機能に必要と思われる残基は変化していない。このことは、この遺伝子が ATPase の構成成分として機能している可能性を示唆する。 α サブユニットには、アミノ末端に 20aa、カルボキシル末端に 60aa 程度の他の α には見られない余計なペプチド鎖が付加している（図17、パネルA）。

A: alpha subunits

maize	chl	10	20	30	40	50	60	70
maize	mit	- - - - -	MATLRVDEINKILRERIEQYDYNRKVGIEENIGRVVQVGBUARIIGLGE					
bovine	mit	- - - - -	-MEFSPEAELTTLEESPMINFYTNLKVDIEGRVVSCCDGIARVYGLME					
<i>E. coli</i>		- - - - -	-EKTGTAEVSSILERILGADTSVOLEETGRVLSIGDGIARVHGLRN					
<i>M. barkeri</i>		MRLIPNQSPHKFPVNWETKSILKDVFDKVFTIEHQV - -	-FE SVTPT-LTPKEVGRILTVSTGIANVSGLPT					
maize	chl	80	90	100	110	120	130	140
maize	mit	IMSGELVEFAEGTRGIALNLSEKNVGIVLMGDGLMLOEGSFVKATGRIAOIPVSEAYLGRVINALAKPIC						
bovine	mit	IQAGEMVEFASGVKGIAUNLNEENVGIVVFGSDTAIKEGDLVVKRTGSIVDVPAGKAMLGRVVDALGPVID						
<i>E. coli</i>		-VQAEEMVEFSSGLKGMSLNLPDNVGIVVFGNDKLIKEDGDIVKRTGAIVDVPVGEEELLGRVVDALGNайд						
<i>M. barkeri</i>		OMOGENMISLPGNRYAIALNLERDSVGAVVMMOPVADLAEGGMKVKGTRILEVPVGRGLLGRVNTLGPAPID						
maize	chl	150	160	170	180	190	200	210
maize	mit	GRGEIVASESRLIESPAPGIIISRRSVYEPLOTGLIAIDSMPIGRGCRELIIGDROTGTATAVATDTILNO						
bovine	mit	GK GALSDHERRRVENKAPGIIERKSVPHEPMOTGLKAVDSLVPIGRGCRELIIGDROTGTATAVATDTILNO						
<i>E. coli</i>		GKGPIGSKARRRVGLKAPGIIIPRISVREPKTGIAVADSLVPIGRGCRELIIGDROTGTATAVATDTILNO						
<i>M. barkeri</i>		GYGPIDHDHGFSAVEATAPGVIERSVDPVOTGYKAVADSLVPIGRGCRELIIGDROTGTATAVATDTILNO						
maize	chl	220	230	240	250	260	270	280
maize	mit	R-----GODVICVYVAIGORASSVAQVVTTFHEEGAMEYTIVVAEFLADSPATLQYALPYTGAALAE						
bovine	mit	KOMNSRGTNESETLYCVYVAIGOKRSTVAOLVOILSEANALEYSMUVAAATASDPAFLAPYSGCAMGE						
<i>E. coli</i>		BRENDD-GTDEKKLYCTYVAIGOKRSTVAOLVKRLTDADAMKYTTIVVSATASDPAFLAPYSGCSMGE						
<i>M. barkeri</i>		R-----DSGIKCHYVAIGOKASTISNVVRKLEEHGALANTIVVVATASESAALOYLARMPVALMGERNF-----NVLCVYCAIGQRASAVARAVANLREGAMDTIVVTEGNDPSGLIYITPYAATSIAE						
maize	chl	290	300	310	320	330	340	350
maize	mit	YFMYPERHHTIYDDLSKQACAYRQMSLLRRPPGREAYLGDFVYLHSRLERAALKNSLLGE-----						
bovine	mit	YFRDNGMHALIIYDDLSKQAVAYRQMSLLRRPPGREAYPGDVFYLHSRLERAALKNSLLGE-----						
<i>E. coli</i>		YFKURGEDALIYDDLSKQAVAYRQMSLLRRPPGREAYPGDVFYLHSRLERAALKNSLLGE-----						
<i>M. barkeri</i>		YFMESGRDVLIYVDDLTNHARAYRELSSLRPPGREAYPGDIFYIHSRLLERSHLKEFFGG-----						
maize	chl	360	370	380	390	400	410	420
maize	mit	-----GSMTALPIVETOSGOVSAYIPTNVISITDQGIFLISADLFNAGIRPAINVGISVSRVGSAAQIKAMK						
bovine	mit	-----GSLTALPIVETOAGDVSAVYIPTNVISITDQGICLTELTYRPGIRPAINVGLSVSRVGSAADLKM						
<i>E. coli</i>		-----GSLTALPIVETQAGDVSAVYIPTNVISITDQGIFLTELTYRPGIRPAINVGLSVSRVGSAAQTRAMK						
<i>M. barkeri</i>		-----GSLTALPIVETOAGDVSAFVPTNVISITDQGIFLETNLFNAGIRPAVNPGISVSRVGCAAQIKIMK						
maize	chl	430	440	450	460	470	480	490
maize	mit	QVAGKSKLELAQFAELQAFASALDKTSQNQLARGRRIRELLIQSQSNPLPVVEQVATIYIICTRYLD						
bovine	mit	DVCGSSKLLELAQYREVAFAAOFGSDLDAATOALLNRGARLTENVKOPOYEPLPTEKQIVVIYAANNGECD						
<i>E. coli</i>		QVAGTMKLELAQYREVAFAAQFGSDLDAATQQLLSRGRVLTTELLKGQYSPMAIEECQVAVIYAGVRGYLD						
<i>M. barkeri</i>		KLSGGIRTALAQYRELAASFQAFASDLDDATRNLQDHGQKVITELLKOKOYAPMSVAQQSLVLFAAERGYLA						
maize	chl	500	510	520	530	540	550	560
maize	mit	SLEIEQVKKFLDELRLKHLKDTKPQFOQEIIISSSKFTTEQAETLLEKAIQEELERFLSQEQT-----						
bovine	mit	RMPIDRISQYEKNIESTI---NPELLKSFLEKGGLTNERKMEPDASLKESENALNL-----						
<i>E. coli</i>		KLEPSITKFEENAFLSHVTSOHOALLSKIRTDGKISE--ESDAKLKEEVTFNFLAGEA-----						
<i>M. barkeri</i>		DVELSKIGSEEALLAYVDRDHAPLMQEINQTYGYNDEIEGKLNQILDSEFKATQSW-----						
maize	chl	570	580	590	600	610		
maize	mit	-----						
bovine	mit	-----						
<i>E. coli</i>		-----						
<i>M. barkeri</i>		NAESETOTEDKSESEAKTEEKVAKSETOTKGKPETETKTEKEPEFGTRTEEK						

(図17) 次ページへ続く

B: beta subunits

maize	chl	MR TNPTTSRPGISTIEEKSVGRIDQIICPVL	10	PPGKLPIIYNALIVKSRDTADKOINVTQEVOOLLG	20		30	40	50	60	70
maize	mit	----DEFTGAGAT-----		-GIOVCOVIGAVV DVRFDLEG-LPPFILTALEV		---	DNNIRLVLVEA	AQHLLG			
bovine	mit	---SPSPKAGATT-----		-GRIVAVIGAVV DVQFDEG-LPPFILNALEVQ		---	GRET	RLVLEVAQHLLG			
<i>E. coli</i>		-----MAI-----		-GKIVUVIGAVV DVEFPQDAVPRVYDALEVO		---	GNERL	VLEVOOOLG			
<i>M. barkeri</i>		---MVDRSRLNL-----		-GKVISVRSITDVLF-EKHLPPVYTILRA		---	GRESQIAIEVLTQD				
maize	chl	NNRVRVAVAMSA	80	TEGLMRGMEVIDTGTPLSVPVG	90	GATLGRIENVLGEPIDLGPVDTSATFP	100	PIHRSAPAFI	110	120	130
maize	mit	ENMVRTIAMDG	140	TGEGLVRGQRVNTGSPITVPVGRATLGR	150	PIINVIGEPIDEKGDIKTNHF	160	FLPIHREAPAFV	170	180	190
bovine	mit	ESTVRRTIAMDG	200	TEGLGVRGQKVLDGAPIRIPVGPE	210	GEPIDERGPIKTKQFAAIHAEAPEFV	220		230	240	250
<i>E. coli</i>		GGIVRTIAMGSSDGL	260	ETLGRIMNVLGEPVDMKGEGEIGEEERWAIHRAAPS	270	YEAHAEAPEFV	280		290	300	310
<i>M. barkeri</i>		AHHVRGIA	320	LTPTEGLARGM	330	EDTGGPLKAPVGRE	340	ILSRMFDVFGNTIDRRKPPSDIOWRS	350	THOTPPPLM	360
maize	chl	ELDTIKLSIFETGIKV	370	VDLAPYR	380	RGGKIGLFFGGAGVGKTVLIMELINNIAKA	390	AHGGGSVFG	400	GVGERTREGN	410
maize	mit	EOATEEQDIL	420	TDLLAPYORGGKIGLFFGGAGVGKTVLIMELINNVA	430	KAHGGGSVFG	440	AGVGERTREGN	450		460
bovine	mit	EMSVEOEIL	470	TDLLAPYAKGKIGLFFGGAGVGKTVLIMELINNVA	480	KAHGGGSVFG	490	AGVGERTREGN	500		510
<i>E. coli</i>		ELNSQELLETG	520	TDLLMCPFAKGKIGLFFGGAGVGKTVLNK	530	MELTRNIAIEHS	540	GGGSVFG	550	AGVGERTREGN	560
<i>M. barkeri</i>		RRSTTSEI	570	TDLLMCPFAKGKIGLFFGGAGVGKTVLNK	580	MELTRNIAIEHS	590	GGGSVFG	600	AGVGERTREGN	610
maize	chl	DLYMEMKE	630	SGVIZN-EKNIEES	640	VALVYQGMNEPPGARMRVGLTALTMAEYFRD	650	VNKODVLLFIONIFRF	660	V	670
maize	mit	DLYREMIE	680	SGVIZNDKQSES	690	KALVYQGMNEPPGARARVGLTGLTVAEYFRD	700	AEGQDVLLFTDNIFRF	710		720
bovine	mit	DLYHEMIE	730	SGVIZNLD-	740	-ATSMVALVYQGMNEPPGARARVALTGLTVAEYFRD	750	EGQDVLLFTDNIFRF	760		770
<i>E. coli</i>		DLYHEMIDSNV	780	ID-----KVSLVYQGMNEPPG	790	PNRLRVVALTGLTMAEKFRD-EGRDVLLF	800	YDNIFYRT	810		820
<i>M. barkeri</i>		ELYRDMNDAGV	830	PL-----PNTVMWFG	840	GMNEPPGARFRVGHALTMAEYFRDDER	850	RDVLLLIONIFRF	860		870
maize	chl	OAGSEVSALLG	890	RHPSAVGYOPTL	900	STENGSLQERITS	910	TKKGSI	920	TS10AVYV	930
maize	mit	QANSFVSALLG	940	PSAVGYOPTL	950	TDLQGLQERITT	960	TKKGSI	970	PADDLT	980
bovine	mit	OAGSEVSALLG	990	GRIPSAVGYOPTL	1000	TDLQGLQERITT	1010	TKKGSI	1020	DPA	1030
<i>E. coli</i>		LAGTEVSALLG	1040	GRIMPSAVGYOPTLA	1050	TDLQGLQERITT	1060	TKKGSI	1070	PATT	1080
<i>M. barkeri</i>		DAGSEVSGL	1090	GRIMPSAVGYOPTLA	1100	TDLQGLQERITT	1110	TKKGSI	1120	FAHL	1130
maize	chl	JVLSRGLASKG	1150	IYPAVDPLDSTS	1160	TDLQGLQERITT	1170	TKKGSI	1180	TDAGAIMS	1190
maize	mit	JVLSRGLASKG	1200	IYPAVDPLDSTS	1210	TDLQGLQERITT	1220	TKKGSI	1230	TDAGAIMS	1240
bovine	mit	JVLSRGLASKG	1250	IYPAVDPLDSTS	1260	TDLQGLQERITT	1270	TKKGSI	1280	TDAGAIMS	1290
<i>E. coli</i>		JVLSRGLASKG	1300	IYPAVDPLDSTS	1310	TDLQGLQERITT	1320	TKKGSI	1330	TDAGAIMS	1340
<i>M. barkeri</i>		JVLSRGLASKG	1350	IYPAVDPLDSTS	1360	TDLQGLQERITT	1370	TKKGSI	1380	TDAGAIMS	1390
maize	chl	TVARARKIERFLS	1400	OPFFVAEVFTGSPG	1410	YVGLAETIRGFOLI	1420	LSGEIDGLPEOAFYLVG	1430	NIDEAS	1440
maize	mit	TVARARKIERFLS	1450	OPFFVAEVFTGSPG	1460	YVGLAETIRGFOLI	1470	LSGEIDGLPEOAFYLVG	1480	TRAIN	1490
bovine	mit	TVARARKIERFLS	1500	OPFFVAEVFTGSPG	1510	YVGLAETIRGFOLI	1520	LSGEIDGLPEOAFYLVG	1530	EEVIAKAEK	1540
<i>E. coli</i>		TVARARKIERFLS	1550	OPFFVAEVFTGSPG	1560	YVGLAETIRGFOLI	1570	LSGEIDGLPEOAFYLVG	1580	EEVIAKAEK	1590
<i>M. barkeri</i>		TVARARKIERFLS	1600	OPFFVAEVFTGSPG	1610	YVGLAETIRGFOLI	1620	LSGEIDGLPEOAFYLVG	1630	EEVIAKAEK	1640
maize	chl	LEEE	1650	SKLKK	1660		1670		1680	1690	1700
maize	mit	TAKE	1710	SAS	1720		1730		1740	1750	1760
bovine	mit	TAEE	1770	H	1780		1790		1800	1810	1820
<i>E. coli</i>		TAEE	1830	H	1840		1850		1860	1870	1880
<i>M. barkeri</i>		TAEE	1890	H	1900		1910		1920	1930	1940

(図17) *M. barkeri* AtpA および AtpI の一般的な F タイプの α (パネルA)、 β (パネルB) サブユニットに対する相同性。chl は葉緑体 F タイプ、mit はミトコンドリアド タイプを示す。- は適正な類似性を保つために挿入されたギャップ。そのサイトで保存されているアミノ酸を黒く塗りつぶしてある。

3.2.5 δ サブユニット遺伝子の欠落

M. barkeri の atp オペロンには δ サブユニットをコードすると思われる遺伝子が見つからない。 δ サブユニットをコードする遺伝子は、通常、b および α サブユニットをコードする遺伝子の間隙に位置する。*M. barkeri* では、この領域に読み枠は見つからず、代わりに AtpF のカルボキシル末端に付加した長いペプチド鎖のコードが続いている。また、最近、立体構造上 b と δ は相互作用をしている可能性が高いとの報告がある (Sawada et al., 1997)。以上の状況から、AtpF カルボキシル末端の付加配列がなんらかの理由で b サブユニットに融合してしまった δ サブユニットではないかという見方は、到達しやすい考え方ではある。しかしながら、アミノ酸配列の類似性検索の結果や疎水性/親水性プロットのいずれもそうした事実は示唆しない (図18)。

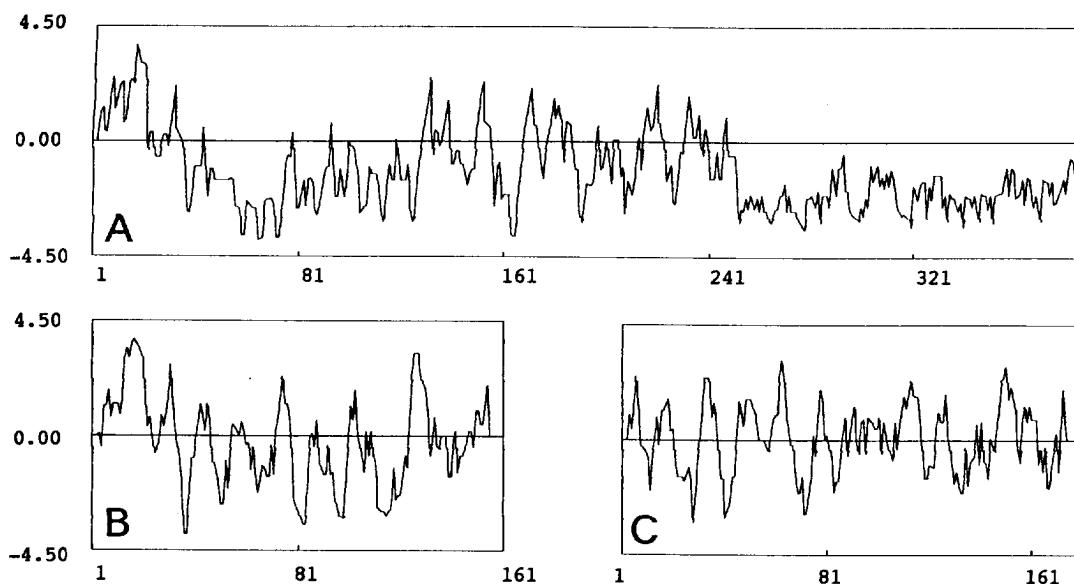


図18) *M. barkeri* AtpF (パネルA) と *E. coli* b (パネルB) および δ (パネルC) サブユニットの疎水性/親水性プロット。bサブユニットはアミノ末端に膜貫通部位と思われる特に強い疎水性領域を持つ。 δ は水可溶性タンパク質であるが全体に疎水性領域を有する。仮に AtpF のカルボキシル末端に δ サブユニットが融合されてしまった場合、 δ の疎水性が見られるはずである。しかし、この領域は逆に親水性になっている。プロット、プラスが疎水性、マイナスが親水性。インデックスは Kyte と Doolittle のものを使用。

3.2.6 γ サブユニット遺伝子の二分化

AtpG (149aa) は、*Arabidopsis thaliana* の葉緑体 F タイプの γ と 23% 程度の相同性を持つ。しかし、一般的な γ サブユニットの長さ（約300aa）に比べて、これは半分程度の長さでしかない。一方、AtpJ (186aa) は、当初、類似性検索には適当な候補が見つからなかったが、検索システムを変えると *Bacillus subtilis* の γ のカルボキシル末端と 31% の相同性を持つことが分かった。この情報を総合すると、AtpG は γ のアミノ末端側の半分、AtpJ はカルボキシル末端側半分をコードすることになる。

これら 2 つの遺伝子の読み替りのオーバーラップは 110 bp と異常に長い（図16、B）。また、GC 含有量の少ない *M. barkeri* には珍しく C の連続する配列も含まれるため、シークエンス反応のミスという側面も考慮し、反応系を変えて再三配列決定を行なったが、読み間違いは発見できなかった。

Walker らの報告したウシ心筋ミトコンドリア F₁ 部分の立体構造[Abrahams *et al.*, 1994]では、 γ サブユニットは、アミノ末端とカルボキシル末端で構成されるコイルドコイル構造により、 $\alpha_3\beta_3$ の構成する球状構造の中心を貫く軸のような構造を形成することが分かっている。仮に、*M. barkeri* の γ がアミノ末端とカルボキシル末端とで別にコードされたタンパク質であっても、こうした構造をとることは十分可能であり、機能上の障害にはならないと考えられる。

3.3 考察

3.3.1 *M. barkeri* の F タイプ遺伝子は機能しているのか？

M. barkeri の atp オペロンは、明らかに F タイプの特徴を持つ遺伝子ではあるが、個々の遺伝子産物の予想される性質は、従来報告されている F タイプのものと大きく異なる点が多い。もし、この遺伝子が発現し機能しているとすれば、F タイプの構造に関する研究に大きなインパクトを与えることは間違いないだろう。この意義をふまえて、*M. barkeri* における atp オペロン発現の証拠を得るべく、ノザンプロット、ウエスタンプロット解析などを行なった。しかし、残念ながらその痕跡を検出することはできなかった。

M. barkeri と同様に、V/F 両タイプの遺伝子を持つ *E. hirae* は、低い pH の際には F タイプの ATP 水解と供役したプロトン輸送で膜電位を作り出し、高い pH もしくは高いナトリウムイオン濃度の環境下では V タイプを用いたナトリウムイオン輸送機能を発現することが知られている [Kakinuma and Igarashi, 1989]。この際、V タイプの活性はその発現量により調節されている。*M. barkeri* の F タイプの発現量も、同じ様な仕組みで pH やナトリウムイオン濃度によって調節されている可能性があると考え、高低 pH (7.5~4.8) 、高低ナトリウムイオン濃度 (400~11mM) の培地を特別に調製し、*M. barkeri* の培養を試みたが、得られた菌体を用いたノザン/ウエスタンプロット解析の結果には反映されなかった。

3.3.2 古細菌における F タイプの分布

その発現および機能の確認はできなかったが、*M. barkeri* に F タイプ遺伝子が存在することは明らかである。これは、遺伝子レベルでは少なくとも、F タイプの分布は真正細菌に限られてはいないことを示すのに十分な証拠となりうるが、古細菌において、どの程度 F タイプの遺伝子が分布しているかどうかは本研究では明らかにできない。同じ属の *M. mazei* ではサザンプロット解析の結果、F タイプ存在の可能性は低いという報告がある。また、全ゲノム解析の終わった、*M. jannaschii*、*M. thermoautotrophicum*、*A. fulgidus* を対象に提供されている検索システムでも、F タイプの類似性遺伝子は見つからないことから、*M. barkeri* の F タイプ遺伝子は古細菌の中でもかなり珍しい存在である可能性が高い。

第4章 古細菌におけるFタイプATPase遺伝子の存在とバクテリア進化との関係

FタイプもしくはVタイプにおいて見られる主要2サブユニットの類似性の高さは、それらが共通の始原遺伝子から分化したものであることを示唆する。また、FタイプとVタイプ自身も互いに機能や構成成分などにおいて高い類似性を持ち、やはり共通の始原タイプから派生したものであることが予測可能だ。バクテリアから高等生物まで一貫して高度な保存性を保ちつつ、これほどまでに広範に分布するタンパク質は他にはないため、ATPaseはrRNAなどと並ぶ生物進化のマーカ遺伝子としてよく用いられてきた。新しく得られた *M. barkeri* Fタイプの情報も含め、改めてF/Vタイプの α/β サブユニットの一次配列を使って分子進化系統樹を作成した結果、Fタイプの系統樹に真正細菌と古細菌の分岐を示すと思われる新しい枝を継ぎ足すことができたので報告する。

4.1 方法

4.1.1 塩基配列の収集

GenBankデータベースを通じて、ATPase関連の登録された遺伝子のうち、 α/β サブユニットの遺伝子を含む塩基配列を引用した。得られた塩基配列と付加された読み枠に関する情報をもとに、想定されるアミノ酸配列に翻訳し情報として用いた。真核生物については、ミトコンドリア、葉緑体（植物の場合）のFタイプ、空胞系のVタイプの遺伝子が揃って報告されているものを探した。使用した遺伝子を保持する生物の分類は表04。[*Halobacterium salinarium*, GenBank HSATPABA; *Haloferax volcanii*, HVATPAB; *Methanococcus jannaschii*, MJU67477; *Methanosarcina barkeri*, MESATPAB; *Methanosarcina mazeii*, MMU47274; *Sulfolobus acidocaldarius*, SSOATPMA; *Enterococcus hirae*, ENENTP; *Thermus thermophilus*, D63799; *Bos taurus*, BOVATPASEV, BOVATPF1A; *Drosophila melanogaster*, DMU19742; *Hordeum vulgare*, HVU36939; *Neurospora crassa*, NEUVMA1A, NEUATPB; *Saccharomyces cerevisiae*, YSCHATP, SCU46215; *Trypanosoma congolense*, TC-VACATPA; *Zea mays*, ZMU36436, CHZMXX, ZMATP2MT; *Cyanophora paradoxa*, CPU30821; *Marchantia polymorpha*, CHMPXX; *Nicotiana tabacum*, CHNTATPBG, NPATP21; *Oryza sativa*, RICCPATP, RICATPB; *Pinus thunbergiana*, PINCPTRPG; *Porphyra purpurea*, PPU38804; *Spinacia oleracea*, SOU23082; *Triticum aestivum*, WHTCPATPB; *Hevea brasiliensis*, HBATPB; *Rattus norvegicus*, RATMTATPSA]

バクテリア

Hsa	<i>Halobacterium salinarium</i>
Hvo	<i>Haloferax volcanii</i>
Mja	<i>Methanococcus jannaschii</i>
Mba	<i>Methanosarcina barkeri</i>
Mma	<i>Methanosarcina mazeii</i>
Sac	<i>Sulfolobus acidocaldarius</i>
Bsu	<i>Bacillus subtilis</i>
Ehi	<i>Enterococcus hirae</i>
Eco	<i>Escherichia coli</i>
Pmo	<i>Propionigenium modestum</i>
Rbl	<i>Rhodopseudomonas blastica</i>
Sli	<i>Streptomyces lividans</i>
Syn	<i>Synechococcus</i> sp.
Tth	<i>Thermus thermophilus</i>

真核生物

Cre	<i>Chlamydomonas reinhardtii</i>	コナミドリムシ
Dme	<i>Drosophila melanogaster</i>	ショウジョウバエ
Sce	<i>Saccharomyces cerevisiae</i>	酵母
Tco	<i>Trypanosoma congoense</i>	トリパノソーマ（寄生虫）
Bta	<i>Bos taurus</i>	ウシ
Rno	<i>Rattus norvegicus</i>	ラット
Spu	<i>Strongylocentrotus purpuratus</i>	ウニ
Ncr	<i>Neurospora crassa</i>	アカパンカビ
Cpa	<i>Cyanophora paradoxa</i>	藻類
Mpo	<i>Marchantia polymorpha</i>	コケ
Ppu	<i>Porphyra purpurea</i>	紅藻類
Hbr	<i>Hevea brasiliensis</i>	ゴム
Hvu	<i>Hordeum vulgare</i>	大麦
Nta	<i>Nicotiana tabacum</i>	タバコ
Osa	<i>Oryza sativa</i>	コメ
Pth	<i>Pinus thunbergiana</i>	マツ
Psa	<i>Pisum sativum</i>	エンドウ
Sol	<i>Spinacia oleracea</i>	ホウレンソウ
Tae	<i>Triticum aestivum</i>	コムギ
Zma	<i>Zea mays</i>	トウモロコシ

〈表04〉

本論文中では由来する生物の学名を3文字に省略したものを使用する。また、それぞれの一次配列を区別するために、1文字目（m、c、v、e、a）で順にミトコンドリア、葉緑体、空胞系、真正細菌、古細菌の区別を、2文字目（f、v）でF/Vタイプの区別を、3文字目（c、r）で触媒サブユニット（Fタイプはβ、Vタイプはα）か非触媒サブユニット（Fタイプではα、Vタイプではβ）の区別を付ける記号をアンダーラインを隔てて先頭に付した。たとえば、古細菌に属する *M. barkeri* のFタイプ触媒サブユニットはafc_Mbaと表記する。

4.1.2 アライメント

GCGの提供するアライメントパッケージ（pileup）を利用し、より多くのアミノ酸残基サイトにおいて最適の類似性を示すように並べた。ギャップ作成に関するパラメータはデフォルト設定を使用（gap creation penalty, 3.00; gap extension penalty, 0.10）。

4.1.3 系統樹の作成

PHYLIP (Phylogeny Inference Package; Washington 大学, J. Felsenstein 提供) を使用した。アミノ酸差算出には ProtDist、系統樹作成には Neighbor、系統樹の描画には Drawtree を使用した。アミノ酸置換行列は Dayhoff らの PAM インデックスを採用。ブツストラップ用のアミノ酸配列は Seqboot を用い 100 セット作成した。アミノ酸差算出、系統樹描画の後、コンセンサス樹形の抽出には Consense を用いた。

4.2 結果

4.2.1 アライメントと相同性マーク

入手した一次構造を一挙に並べ、GCG パッケージによりアライメントを取った（付録C）。計算には挿入配列や欠失の影響を受けていない保存配列を用いる。挿入配列や欠失を受けた領域は手作業で除いた。保存領域を見つけやすくするために、*B. subtilis* の F タイプ触媒サブユニットと相同的なアミノ酸について黒く塗りつぶしマークした。こうして抽出されたサイトの数は約 380 であった。

4.2.2 系統樹の作成

真核生物では、植物、動物、菌類を、真正細菌では Eco、Bsu の他に、V タイプの報告されている Tth、Ehi、ミトコンドリアおよび葉緑体のもとになった生物と比較的近縁と思われる Rbl、Syn を加えた。古細菌は Sac、Hsa、Mba を使用。Mba は本研究で明らかになった F タイプも加えた。以上の生物分類において、それぞれの F タイプ、V タイプの触媒/非触媒サブユニットがどのような系統樹を形成するのかを確認した（図19）。

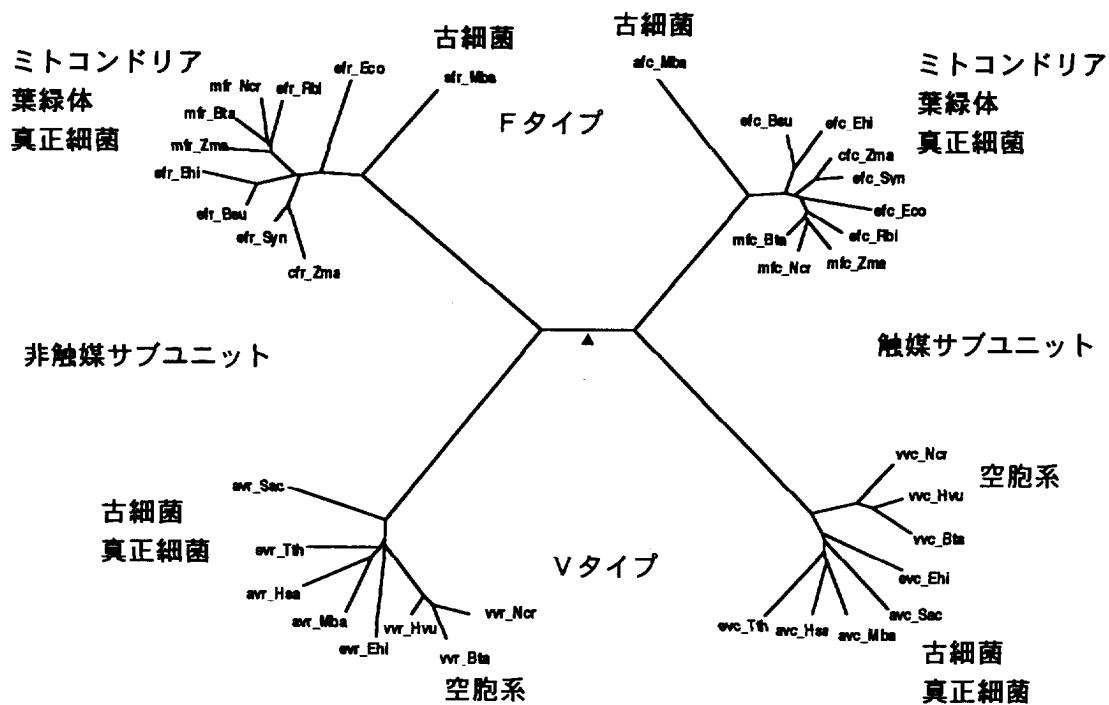


図19) 触媒/非触媒サブユニット、F/V タイプの系統樹。

4.2.3 系統樹の統計的検定

得られた系統樹の信頼性を調べるのには一般にブーツストラップ法が用いられる[Nei, 1987]。ブーツストラップ法というのは m 個のアミノ酸サイトからなる複数の配列がある場合、この中から m 個のサイトを重複を許して再抽出し、この新しいデータからまた系統樹を作成するものである。重複を許す再抽出により、サイト数を無限大と仮定した場合、その集団から m 個のサイトを無作為に抽出したのと

同じことになる。もし、 m が大きければ、再抽出後のデータから得られる系統樹と元の系統樹は、ほぼ同じ形になる。しかし m が小さいときは異なった樹形になる可能性がある。したがって、こうした再抽出を何度も行なえば、元のデータで得られた系統樹の信頼度を調べることができる。もし、特定の分岐パターンが 95% 以上の頻度で得られれば、それはかなり信頼できると考えられる。図20 に 100 回のブーツストラップにより得られた系統樹を示す。数字はそれぞれの分岐点の出現回数を示している。

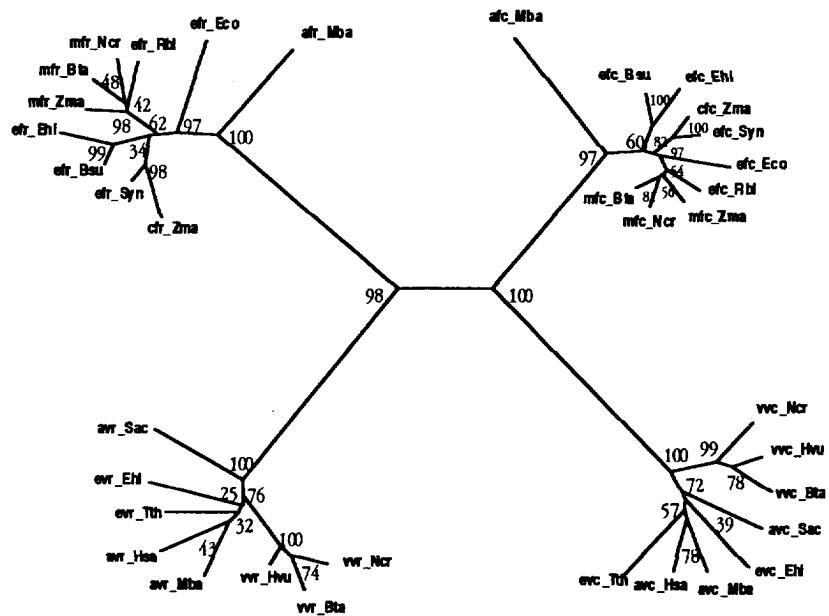


図20) ブーツストラップ法により得られた系統樹。数字は 100 パターンの重複を許す無作為抽出結果から導き出された樹形における各分岐の出現回数を示す。100 に近いほど出現回数が多く、誤差の少ない確からしい分岐であることを示す。

本研究においてブーツストラップにより得られたコンセンサス樹形は、大まかには元のデータで得られた樹形と同一であり、触媒/非触媒サブユニットの分岐により形成される樹形や、V/Fへの分岐により形成される樹形についてはブートストラップ値も 100 に非常に近く信頼できると思われる。一方、細かな点、たとえば図の右上に位置する F タイプ触媒サブユニットの樹形で、ミトコンドリアのもの (*mfc_****; *** は Zma、Ncr、Bta) が形成する樹形や、右下の V タイプ触媒サブユニットの樹形で、古細菌の Sac と真正細菌の Ehi の枝の位置が入れ替わる、などといった元のデータから得られる系統樹と異なる点が観察される。また、こうした誤差を生じる分岐のブーツストラップ値は一般に 40~80 と低い値を示しているため信頼できない。

4.3 考察

4.3.1 触媒/非触媒サブユニット、F/Vタイプの分化

得られた樹形より、触媒/非触媒サブユニット、F/Vタイプの分化の順序を予測することができる。樹形の対称性より、図19 で▲で示した場所にこの系統樹の根を仮定すると、ATPase はまず、触媒サブユニットから非触媒サブユニットが遺伝子重複により生じた後、F タイプと V タイプにやはり遺伝子重複により分化したことが示唆される。当然のことであるが、このことは、同様の手法により系統樹を作

成したり、アミノ酸配列の類似性に着目した過去の知見とは矛盾しない[Gorgarten et al., 1989; Nelson, 1992]。

4.3.2 新たに書き加えられた古細菌のFタイプの枝と真正細菌/古細菌の分化

本研究により、Fタイプの系統樹（図20の上半分）に新たに2本の枝を継ぎ足すことができた。触媒サブユニットと非触媒サブユニットのそれぞれ根に近い側にあるafc_Mba、afr_Mbaとラベルした枝がそれである。ブーツストラップ値もそれぞれ97、100と十分大きく、第3章の他の種の α および β サブユニットとの相同性比較からも直感的に推察されたように、古細菌のFタイプ遺伝子産物が、他のFタイプ遺伝子産物とは一線を画す存在であることが系統樹からも裏付けられた。

M. barkeri のFタイプ遺伝子が古細菌のFタイプを代表することが可能な存在かどうかは、1種しか発見できていない現段階では明言できない。この新たに書き加えられた*M. barkeri* Fタイプの枝に、今後、発見されるであろう他の古細菌のFタイプの遺伝子産物の枝が継ぎ足されるならば、このFタイプの系統樹を用いて、真正細菌と古細菌の分岐の時期を推察することが可能になるだろう。つまり、従来考えられていたように[Gorgarten et al., 1989; Iwabe, N. et al., 1989]、F/Vの分岐と真正細菌/古細菌の分岐時期は一致するのではなく、F/Vの分岐が起こってからかなり経過した後に真正細菌と古細菌の分化が起こったのであることを示すことが可能になる。

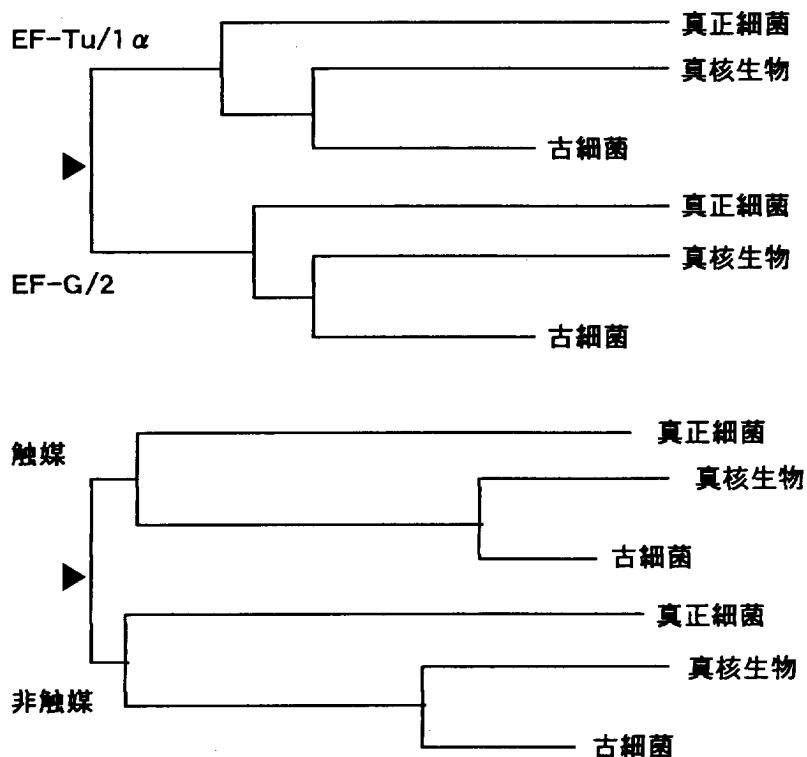
繰り返しになるが、このことを示すには、*M. barkeri* 以外の古細菌、できれば高度好塩菌や好酸好熱古細菌からFタイプの遺伝子を発見する必要がある。

4.3.2 系統樹の根

Woeseらが古細菌を提唱して以来、古細菌/真正細菌/真核生物を結ぶ系統樹のどの部分に根、つまり始原生物からの分岐点を設けるかについて議論が重ねられてきた。現存生物の分子の比較から、無根の系統樹を算出することは可能だが、この系統樹には時間の軸がないため、どの順序で各枝が分岐したのかを推察することはできない。無根の系統樹から有根の系統樹を作るには、比較した生物グループに比べ明らかに古い時期に分岐したことが分かっている別の生物の系統を比較に含める必要がある。しかし、全生物の有根系統樹を描く際に、現存する生物からこの目的に適したものを選ぶことはできない。

宮田らは、複合系統樹法によりその問題を解決した[Iwabe, N., et al., 1989]。古細菌、真正細菌、真核生物の共通の祖先を根とするのではなく、さらに古い時代にさかのぼって根を設定しても同じ結果を導くことができる。つまり、遺伝子重複により生じた双子の遺伝子産物の系統樹、複合系統樹を利用することで派生的に有根系統樹を作成したのである。遺伝子重複は三者の分岐に先行するので、作成された系統樹の対称性より重複時期を決定することができ、しいては、この重複点を境に2つの有根系統樹が自動的に作成される。

複合系統樹から古細菌、真正細菌、真核生物の系統関係を知るために利用できる分子には制約がある。第1に超生物界に共通の遺伝子であること。第2に進化がたどれる程度に保存されていること。第3に、第2の条件を満たしていて、なお統計的計算に耐えられるよう十分な長さの配列であること。以上の3点を満たしている必要がある。このような条件を満たす遺伝子の数は少ない。宮田らは延長因子EF-Tu/1 α とEF-G/2の遺伝子対およびF/V-ATPaseの触媒/非触媒サブユニットの遺伝子対のそれぞれに対して複合系統樹を作成し、Woeseの系統樹に根をつけることに成功した（図21）。



〈図21〉 Iwabe, N., et al., 1989 を改変。

この宮田らの研究成果はたいへんすばらしいものであったが、ただひとつ、F/Vタイプの分岐時期を真正細菌/古細菌の分岐時期と一致させる説（第1章、図02）に乗っ取っていること、つまりFタイプは真正細菌にのみ、Vタイプは古細菌にのみ分布するという前提に立ちATPaseに関する系統樹を作成している点で議論の余地が残されていると考えた。

本研究により、古細菌のFタイプの遺伝子を新しく利用できるようになった。しかし、系統樹からもうかがえるように細胞内小器官のFタイプはすべて真正細菌が細胞内共生により持ち込んだものである可能性が高い。そこでFタイプを用いて、系統樹の根、ひいては細胞内共生において宿主となった真核生物の始原生物を明らかにするためには、宿主由来のFタイプに情報を求めなければならない。しかし、真核生物の細胞膜や空胞系からFタイプは見つかっていない。これは先に述べた複合系統樹に用いる情報の第1条件「普遍的な分布」に反する。残念ながらFタイプで系統樹の根を決定する作業はできない。

一方、本研究における古細菌Fタイプ発見とは直接関係はないが、Vタイプには真正細菌のVタイプの2例を含め、真正細菌、古細菌、真核生物（宿主由来と想定される空砲系）のすべてから情報を得ることができる。比較的高度に保存され、十分な長さを持つこのVタイプの触媒/非触媒サブユニットをコードする遺伝子対は、複合系統樹の作成に向いていると思われる。しかし、今回作成した系統樹では特にVタイプの真正細菌の枝（図19のラベル ev*; * は c または r）はブートストラップ値が小さく、分岐位置に関する誤差が大きいため信頼性に欠け、根の位置を特定するといった議論には耐えられない。

本研究の他にも、Kibak らや Gogarten と Hilario がVタイプを用いて根の決定を行なおうとしているが、いずれの研究においても真正細菌のVタイプの枝は誤差を含んでおり、明確な結論には至っていない。また、両者は誤差を含みながらも本研究で作成した系統樹と同様に真正細菌のVタイプの枝が好塩菌およびメタン産生菌の形成する枝につがれる結果を受けて、*T. thermophilus* や *E. hirae* により提供される真正細菌のVタイプの遺伝子は、遺伝子の水平移動によって古細菌より獲得したものであるという仮説を提唱している[Kibak, H., et al, 1992; Hilario, E. & Gogarten, J. P., 1993]。

他方、宮田は、真正細菌に属し、Vタイプがすでに報告されている *T. thermophilus* と同属の *Thermus aquaticus* YT-1株のリンゴ酸デヒドロゲナーゼが、真正細菌よりむしろ真核生物のものによく似ていること、*T. thermophilus* が古細菌と等しいVタイプを保持していることを併せて、研究途上と前置きしながらも *T. aquaticus* こそが始原生物（細胞内共生説の宿主）であるという仮説を述べている[日本生化学会編、新生化学実験講座(16)分子進化実験法、第4章]（しかし、文献[Iwabe, N., et al., 1989]では、遺伝子の水平移動や、*Thermus* が古細菌であるという可能性にもふれている点は興味深い）。

生物進化のマーカとして、類希な好条件を備えていると思われた ATPase であったが、前述のような理由で、Fタイプ、Vタイプともにマーカとしてはそれほど有用ではないという結論を、少なくとも現時点においては導きざるをえないことは残念である。分子進化研究の立場では、今後の発展として、1) 複数の古細菌からFタイプの遺伝子が発見され、今回Fタイプの系統樹に継ぎ足された *M. barkeri* の枝が、Fタイプにおける古細菌の典型的な枝であることが証明される、2) 多くの真正細菌のVタイプの遺伝子が発見され、誤差の少ないVタイプの系統樹の作成や、現在報告されている真正細菌のVタイプが遺伝子の水平移動で獲得されたものであることを系統樹から証明できるようになる、ことが期待される。

引用文献

- Abrahams, J. P., Leslie, A. G. W., Lutter, R. and Walker, J. E. (1994) *Nature* **370**, 621-8.
- Allmansberger, R., Knaub S. and Klein, A. (1988) *Nucl. Acids Res.* **16**, 7419-36.
- Boone, D. R., Whiteman, W. B. and Rouvière, J. (1993) *Methanogenesis*, New York: Ferry, J. G. 62.
- Boyer, P. D. (1987) *Biochemistry* **26**, 8503-7.
- Boyer, P. D. (1997) *Annu. Rev. Biochem.* **66**, 717-49.
- Cozens, A. L. and Walker, J. E. (1987) *J. Mol. Biol.* **194**, 359-383.
- Duncan, T. M., Parsonage, D. and Senior, A. E. (1986) *FEBS. Lett.* **208**, 1-6.
- Downie, J. A., Gibson, F. and Cox, G. B. (1979) *Annu. Rev. Biochem.* **48**, 103-31.
- Downie, J. A., Cox, G. B., Langman, L., Ash, G., Becker, M. and Gibson, F. (1981) *J. Bacteriol.* **145**, 200-10
- Denda, K., Konishi, J., Oshima, T., Date, T. and Yoshida, M. (1988a) *J. Biol. Chem.* **263**, 6012-15.
- Denda, K., Konishi, J., Oshima, T., Date, T. and Yoshida, M. (1988b) *J. Biol. Chem.* **263**, 17251-17254.
- Denda, K., Konishi, J., Oshima, T., Date, T. and Yoshida, M. (1989) *J. Biol. Chem.* **264**, 7119-21.
- Denda, K., Konishi, J., Hajiro, K., Oshima, T., Date, T. and Yoshida, M. (1990) *J. Biol. Chem.* **265**, 21509-13.
- Dunn, S. D. and Futai, M. (1980) *J. Biol. Chem.* **255**, 113-8.
- Foster, D. L. and Fillngame, R. H. (1982) *J. Biol. Chem.*, **257**, 2009-15.
- Fry, D. C., Kuby, S. A. and Mildvan, A. S. (1986) *Proc. Natl. Acad. Sci. U.S.A.* **83**, 907-11.
- Gorgarten, J. P. Kibak, H., Dittrich, P., Taiz, L., Oshima, T., Konishi, J., Denda, K. and Yoshida, M. (1989) *Proc. Natl. Acad. Sci. U.S.A.* **86**, 6661-5.
- Graber, P. (1987) *Bioelectrochemistry II*, Plenum Publishing Corp., 379-429.
- Higgins, C. F., et al. (1986) *Nature* **323**, 448-450.
- Hilario, E. and Gogarten, J. P. (1993) *BioSystems* **31**, 111-9.
- Hisabori, T., Yoshida, M. and Sakurai, H. (1986) *J. Biochem.* **100**, 663-70.
- Iwabe, N., et al (1989) *Proc. Natl. Acad. Sci. USA* **86**, 9355-59.
- Kakinuma, Y. and Igarashi, K. (1989) *J. Bioenerg. Biomembr.* **21**, 679-92.
- Kakinuma, Y. and Igarashi, K. (1994) *J. Biochem. (Tokyo)* **116**, 1302-8.
- Kibak, H., Taiz, L., Starke, T., Bernasconi, P. and Gogarten, J. P. (1992), *J. Bioenerg. Biomembr.* **24**, 415-24.
- Margulis, L. *Symbiosis in Cell Evolution*, New York: Freeman, W. H.
- Nei, M. (1987) *Molecular Evolutionary Genetics*, New York: Columbia Univ., Press.
- Nelson, N. (1992) *Biochem. Biophys. Acta* **1100**, 109-124.
- Nelson, N. and Taiz, L. (1989) *Trends Biochem. Sci.* **14**, 113-6.
- Sawada, K., Kuroda, N., Watanabe, H., Moritani-Otsuka, C. and Kanazawa, H. (1997) *J. Biol. Chem.* **272**, in press.

- Steinert, K., Kroth-Pancic, P.G. and Bickel-Sandkotter, S. (1995) *Biochim. Biophys. Acta* **1249**, 137-144.
- Sumi, M., Yohda, M., Koga, Y. and Yoshida, M. (1997) *Biochem. Biophys. Res. Commun.* **241**, 427-433.
- Takase, K., Yamato, I. and Kakinuma, Y. (1993) *J. Biol. Chem.* **268**, 11610-6.
- Takase, K., Kakinuma, S., Yamato, I., Konishi, K., Igarashi, K. and Kakinuma, Y. (1994) *J. Biol. Chem.* **269**, 11037-44.
- Tsutsumi, S., Denda, K., Yokoyama, K., Oshima, T., Date, T. and Yoshida, M. (1991) *Biochmin, Biophys. Acta* **1098**, 13-20.
- Walker, J. E., Saraste, M., Runswick, M. J. and Gay, N. J. (1982) *EMBO J.* **1**, 945-51.
- Walker, J. E., Fearnly, I. M., Gay, N. J., Gibson, B. W., Northrop, F. D., Powell, S. J., Runsweick, M. J., Saraste, M. and Tybulewicz, V. L. J. (1985) *J. Mol. Biol.* **184**, 677-701.
- Weber, J. and Seinior, A. E. (1997) *Biochim. Biophys. Acta* **1319**, 19-58.
- Woese, C. R. (1987) *Microbiol. Rev.* **51**, 221-71.
- Yokoyama, K., Oshima, T. and Yoshida, M. (1990) *J. Biol. Chem.* **265**, 21946-50.
- Yokoyama, K., Akanabe, Y., Ishii, N. and Yoshida, M. (1994) *J. Biol. Chem.* **269**, 12248-53.
- Yoshida, M. and Allison, W. S. (1986) *J. Biol. Chem.* **261**, 5714-21.
- Yoshida, M., Sone, N., Hirata, H., Kagawa, Y. and Ui, N. (1979) *J. Biol. Chem.*, **254**, 9525-44.

付録A *M. barkeri* F タイプ ATPase 遺伝子の全塩基配列と予想されるアミノ酸配列

10 20 30 40 50 60 70 80 90 100 110 120 130 140 150
 GATCCGAACGCATCGCAGGGCAGATTCCGTAACCTGTAACTTACCAACCCCTCCGAGTGAGACAGACGGAGCAGTCGTAGTTGTATCTGGATATGGGGCTTTGTGAATACATTCCGAAATGTCAGCTCTGATCTTATAG
 160 170 180 190 200 210 220 230 240 250 260 270 280 290 300
 GCCAGAACAGATTCTCAGGTATCGCTCATGGGCCCGCATATGATACCGCATATGGCTGAACGCTTCAATGGAAACACATCAACGAATGAATGGCTTCAGTGTGGAAAGGTCAAGTGTGGCTGACAAGCATGGTCCGATCT
 310 320 330 340 350 360 370 380 390 400 410 420 430 440 450
 ACAGCCAACTGGTAATGGCCCTCATCAATGGCTACAGGTTGCCAGACGGCTGAAAGTCTGGACTTGACTGGATAACGGTAAATTGTCATCAGACGATTTGGCATGCAAACCTGGGAAACGGCATATGCATTGCTGG
 460 470 480 490 500 510 520 530 540 550 560 570 580 590 600
 AGTACCTGGAAAATGACGATCTGATTCAATTCTCGTGGGACCAATTGATCAATTACAACACTTCAACCTGCTCTGGAGCAGTTCTGAGATAACACGACAATATGTTCTGGAGCATGAGATCTGGCATCTGGAGTTGG
 610 620 630 640 650 660 670 680 690 700 710 720 730 740 750
 AACCTTATTCTGAAAATCCGTTGGAGATGTAATAGAGATAAAAGGAACCGGCTTGTGGAGATGCTGCAAAGTGGGGCACCTTGAAAAGATGTTGAGGATACAGGATGGAGAATACGAATACTGCTTGAAGATAATAATCC
 760 770 780 790 800 810 820 830 840 850 860 870 880 890 900
 CTTCAGGGTTCGACAACAGTTACTGTGAGGCAACAGGTGCCGATGACCTTAATGAGAGCAAAGATTCTGTGGTTAACGAAAAGTGCAGGACAAAACGGTACTGCTACTGTTCCAGAAAGGTGTTCCAGGAACAT
 910 920 930 940 950 960 970 980 990 1000 1010 1020 1030 1040 1050
 ATAAGATCAGATCGATGGAAAGGAAAGGAACTGGCTTAATGTCAAAATCAAGATCACAGCTGTGCAAAGGGTAGATTGAGAAGTTCAGAAGGAAATCTCAACTATGAATAATAACTAAGTCTATAACGGGAGCAATTAAAGTAAAGTCG
 1060 1070 1080 1090 1100 1110 1120 1130 1140 1150 1160 1170 1180 1190 1200
 GAGGCATTGAAAAGCAGATTGAAATTAAACCTTCAGAGAGATAAAGCTGAGAAAGCTTGGACAAAGGCTTCTGGATAATTGATCATCATATAATGCTTCATAAAAATCAATGCACTGTTGAAATT
 1210 1220 1230 1240 1250 1260 1270 1280 1290 1300 1310 1320 1330 1340 1350
 TCATCTTACATAGAGAATGGTGAAGAGCCAAAATTATATTGATAGCATTCTTACGCTAACCGATAACGATCTGCTTCTGCTTAAAGAATAACACGTTATTGACATCCCTTAAAGAAAACACAGTATTGAGACATCCCTGGAGACACAGTCG
 1360 1370 1380 1390 1400 1410 1420 1430 1440 1450 1460 1470 1480 1490 1500
 TATGTTTACAAAAAAGGAGTTCTGGTACGAAAATTCTAGGATTGATAATAAAATTATTATAACATTAAACCCCTGAAAATGATTGTTAGGGATATTAAACCTTATGGGACATTTGCTTAAAGAATAACACAGTATTGAGCTCTCT
 1510 1520 1530 1540 1550 1560 1570 1580 1590 1600 1610 1620 1630 1640 1650
 CGCTAAGGTACATTAAAGAATTATTAAAGGAATTAAAGTACTTAAAGGTAATTAAACGTTAAATTCTTAAACCTTAAACCCCTGAAAATGATTGTTAGGGATATTAAACCTTATGGGACATCCCTGGAGACACAGTCG
 1660 1670 1680 1690 1700 1710 1720 1730 1740 1750 1760 1770 1780 1790 1800
 GAATTGAAAAAGATAAAAGTATTATCTCGCTCTCTGATTAAAAGACTTATTAGATTATTCTAATTTGATTTAGGGATATTAAACCTTATGGGAAATTCTAAATGCAACCTTATGGGAAAGGGAAACAAAGTA
 1810 1820 1830 1840 1850 1860 1870 1880 1890 1900 1910 1920 1930 1940 1950
 CCTCTAGTTTACTGATTGGAGGACCTATATAACATTGAAATTAAACCTGATTGATTGAGCTTCAATGAGACACTATAATGAGTTCTTAAAGACAATAATTGAAAATTGATTCTTCTGGGAAAGGGAAACAAAGTA
 1960 1970 1980 1990 2000 2010 2020 2030 2040 2050 2060 2070 2080 2090 2100
 ATAAAGAATTCACTGGTAAATGGGTTATAAGCACTGAAAGACACTATGAAGACACTATAATGAGTTCTTAAAGACAATAATTGAAAATTGATTCCATTCCATGTTCTTCTGGGACTGTGGCTTGTGA
 2110 2120 2130 2140 2150 2160 2170 2180 2190 2200 2210 2220 2230 2240 (AtpD)
 TGAGGAGATGTAATCTCATATTCTTAGTCTACAAACAAAGCTTGAAGAGGGCACTCCAAACCGAATATAATGCCAGTTGGGGAGCTGTTAAAATAGCAATATAATGCCAGTTGGGGAGCTGTTAAAATGTA
 M V D
 R S R L L N L G K V I S V R G S I V D V L F E K H L P P V Y T L L R A G R E S Q I A I E V L T Q L D
 2260 2270 2280 2290 2300 2310 2320 2330 2340 2350 2360 2370 2380 2390 2400
 TCGAAGTCGGCTCTGAATTCGTAAGGTATCTCGTGGCGGGCACTTGAAGAGGGCACTCCAAACCGAATATAATGCCAGTTGGGGAGCTGTTAAAATAGCAATATAATGCCAGTTGGGGAGCTGTTAAAATGTA
 A H H V R G I A L T P T E G L A R G M A V E D T G G P L K A P V G R E I L S R M F D V F G N T I D R
 2410 2420 2430 2440 2450 2460 2470 2480 2490 2500 2510 2520 2530 2540 2550
 CGCACATCATGTCGCGGAATTGCTGACCCCCACTGAAGGCTTCTGCTGAGGATGGCAGTGGAAAGATAACGGTGCACCTTGAAGGCTGTTGGAGGGAAATCTCTGCAATGTCATGTTGGAAATACCATAGACCG
 R K P P S D I Q W R S I H Q T P P P L M R R S T T S E I F E T G I K A I D V L V P L E R G G K A G L
 2560 2570 2580 2590 2600 2610 2620 2630 2640 2650 2660 2670 2680 2690 2700
 CAGGAAGCCTCCATCTGATATCCAGTGGGCTCAATACATCAACCCCGCACCGTAATGGCTGATCCTACGCTGAAATTCGAGACGGGTACAAAGCTATAGGTTGACGGCTTGGAGGGAAAGCAGGTT
 F G G A G V G K T V L L T E M I H N V V K Q H Q G V S I F C G I G E R C R E G E E L Y R D M K D A G
 2710 2720 2730 2740 2750 2760 2770 2780 2790 2800 2810 2820 2830 2840 2850
 ATTGGAGGAGGGGGTTGGCAAGACAGTGTGCTAACCGAAATGATTCTACGCTAGTCAACACGAGGGAGTGAGCATCTTGTGGAAATAGGCAGGCTGTCAGAGGGAGGAAACTTATCGTATGAAAGACGCTG
 V L P N T V M Y F G Q M N E P P G A R F R V G H T A L T M A E Y F R D D E R R D V L L L I D N I F R
 2860 2870 2880 2890 2900 2910 2920 2930 2940 2950 2960 2970 2980 2990 3000
 VTGCTTCCGAATACGGTTGGTATCGGCGAGATGACGAGGACCCAGGGTGCCTTCTGCTAGGTACAGCGCTGACAAAGCAGAGTACCTTCAAGAGACGATGAACGCGAGATGCTACTGCTTATAGATAACATTTCAG
 V L P N T V M Y F G Q M N E P P G A R F R V G H T A L T M A E Y F R D D E R R D V L L L I D N I F R

3010 3020 3030 3040 3050 3060 3070 3080 3090 3100 3110 3120 3130 3140 3150
 GTTATTTCAGGCTGGTCCGAAGTATCGGTTGATGGGACAGATGCCCTACGCCCTGGGATATCAACCGACATTGGGACTGAATTTCGGAGTTAGAAGAGCGCATATCACCCGATGCCGGAGCCATAATGCTATTCAAGCGGT
 F I Q A G S E V S G L M G Q M P S R L G Y Q P T L G T E L S E L E E R I S T T D A G A I M S I Q A V

 3160 3170 3180 3190 3200 3210 3220 3230 3240 3250 3260 3270 3280 3290 3300
 GATGTGCCCTGCTGACGATTCACTGACCCCTCGCTGTGCATACATTCTCACATCTGCCATCAATTGCTCTCGCAAGAGGGCAAGGCCAGGGCTTTATCCGGTATTGACCCCTTGAGCTAAACTCCAAATGCCACACC
 Y V P A D D F T D P S A V H T F S H L S A S I V L S R K R A S E G L Y P A I D P L Q S N S K M A T P

 3310 3320 3330 3340 3350 3360 3370 3380 3390 3400 3410 3420 3430 3440 3450
 TGGATTGTTGGCAAGGCATTATCTCTGGCCAAGAAATCAGGAGAGCTGGCAAACTCAGAACGTTACAGGATATTATCTCAATGCTTGCGCTGGAGCAGTATGCCCTGGAGGCCAATGTTGAGCTGGTAGCCGGCTGACGGCT
 G I V G E R H Y L L A Q E I R Q T L A Q Y S E L K D I I S M L G L E Q L S P E D R N V V A R A R R L

 3460 3470 3480 3490 3500 3510 3520 3530 3540 3550 3560 3570 3580 3590 3600
 GGGACCTTCTGACGCCAGGGTTTACCCACTGAGCAATTACAGGTAAAGGCAATTCATAGCTTCAAGGCTGACGGCTGCCAGGGCTGGAGGCTATCTGAATGACAATTCAAAGACTACTCTGAAGGTGACCTCTACATGAT
 E R F L T Q P F F T T E Q F T G I K G K S I S L S D A L D G C E R I L N D E F K D Y S E G D L Y M I

 3610 3620 3630 3640 3650 3660 3670 3680 3690 3700 3710 3720 3730 3740 3750
 CGAACAAATTGACGAAGCAATGGCCAAAAAAATCAAGCGGGAAAATCATGAATTCAAGGACTCATGAATTCAAGGACTCATGAATTCAAGGACTCATGAATTCAAGGACTCATGAATTCAAGGACTCATGAATTCAAGGCTATCAAG
 G T I D E A M A K K S S R E K S *
 M N S G L M N S G L M N S G L M N L T I L L P F Q V F A E K K G V S

 3760 3770 3780 3790 3800 3810 3820 3830 3840 3850 3860 3870 3880 3890 3900
 CGCATAGTCGAGAAGGCCGTGGGGCTCATGGGCTCTGGCACACGGACTGTGGTGTGGACTCTGGGCTCTGGAAATTCTCACATGAAACCGAATCAGAAGGGAGGTTATGGCAGTCGATGAAGGCGTACTCATCAAG
 R I V A E G R E G S F G L L P H R L D C V A T L G P G I L T Y E T E S E G E V Y V A V D E G V L I K

 3910 3920 3930 3940 3950 3960 3970 3980 3990 4000 4010 4020 4030 4040 4050
 ATGGTCACTATGACTTGTCTGTACGTGATGCCATTGGGACAGACTTAAGGCAATTGACAGGGCTGTTGAAAAGAGTTTGACCTTAGTGAACCCAGGCAAAATTGCTCTAGTAATGGTAAATTGGAACAGGGCT
 N G Q Y V L V S V R D A I F G T D L S Q L H E A V E K E F L T L D E T E Q K I R S V M V K L E T G L

 (AtpC)
 4060 4070 4080 4090 4100 4110 4120 4130 4140 4150 4160 4170 4180 4190 4200
 ATAGCCGGATTAGCGGAGTTCAAGATGCTGACAAACCATCAAGAACCTTGGAGGAGCAGTCACCTCTGGCCCTCAGTTGGGACAAGGCTGAACGCAAATAAGGGCACACGCTATGAGCAGGACGATCTGGCTGGTT
 I R R L A E F Q N V *
 M S D K P S K K P L E D E S P L A R Q V G T K A E R K I R A Q R H V D R T I W L G L

 4210 4220 4230 4240 4250 4260 4270 4280 4290 4300 4310 4320 4330 4340 4350
 AGGCATGATGGGACTCATAGCTGGCGTAGCAGTCAACTTAATTGGCTGCAATGGGCTCTGGTAGATAAACACTATCCAGAAAGCTTCTGGGACTCACGATGCTGATTGGGCTTATATCGGTTGCTGAATG
 G M M G L I G W S V A V P T L I G A A I G L W L D K H Y P E S F S W A L T M L I I G L I I G C L N A

 (URF2)
 4360 4370 4380 4390 4400 4410 4420 4430 4440 4450 4460 4470 4480 4490 4500
 GTGGCACTGTTAGCTAAGGAGGCCAGGAATGCAAGGAAACAGGGATTACAATGAAATTTGAATCTTCTGCTGATAGCTGGTTCTGCTGGAGCAGTGTCTGGGGCTCTGGGACCGTCAAGAACAGGGCT
 W H W L A K E R Q E M Q E E Q E D Y N E *
 M N E I I L N L F L S L I A G F L L G A V F F G G L W W T V Q K G

 4510 4520 4530 4540 4550 4560 4570 4580 4590 4600 4610 4620 4630 4640 4650
 GACTTTCATCAGAAACCTGCTCTGGTCTGGGAGCCTGTTGACCGCTAGCATGGCTGATAGCTGGGACTTACTTGTCTGCTGATAGCTGGTTCTGCTGGAGCAGTGTCTGGGGCTCTGGGACCGTCAATCAA
 L S S R K P A F W F L G S L L I R L S V I A G F Y F V S D G Y W E R L L I C L F G F F V M R H I I

 (AtpB)
 4660 4670 4680 4690 4700 4710 4720 4730 4740 4750 4760 4770 4780 4790 4800
 TAGTTAGACTCACCAGATTGCCAGAGGAACCCAAATCAATTGACAAAGGAGGCCAGTAATGCACTTGTGCTGAGCTTAATGGCAGTACGACTTATCAACTCAATGCAACGATTGATACTTGGGGCTGATGCTG
 M R L S P D E L I F W Q Y D F I K L N A T I V Y T W G L M L
 V R L T R L P E E D P N Q L T K E A S N A T *

 4810 4820 4830 4840 4850 4860 4870 4880 4890 4900 4910 4920 4930 4940 4950
 GAAATGTTAAATGGTCAAAAATTAACAGCAAACCTCTACTGCTGGACGTTCCCGCTGGAAAATATTCTGAAATTATGTCACAGGCACTCTGGAAACAAATTGAGGACGTTGGCTGGCCCTGGCACCGGAGAAATATCAGG
 V M V I G S K T I T S K L S T G L E R S R W Q N I L E I I V T G I L E Q I E D V G L D Q P R K Y L G

 4960 4970 4980 4990 5000 5010 5020 5030 5040 5050 5060 5070 5080 5090 5100
 TTTTGGGACACTCTCTGTCATTGCTGAGCCAATCTGCATATAATTCCAGGTTGAACCGGCAACAGGATCTCTCAACTACGGCTGCTCTGCAATTGTTGTTGAGCTGTTCACTTCCGGCATGAGAACAA
 F L G T L F L F I A V A N L C I I P G Y E P P T G S L S T T A A L A L C V F V A V P L F G I E E Q

 5110 5120 5130 5140 5150 5160 5170 5180 5190 5200 5210 5220 5230 5240 5250
 GGAAATTAGTAACCTTAAGCTACAGAGGCCACATCATGCTGCCATTAAATATAAGTAAATCTCCGGACGTTGCTCTGCAATCCGGCTCTGCTGATAGCTGGCTCTGCTGATAGCTGGCATCTGGCT
 G I S N Y L K T Y T E P T I M L P F N I I S E I S R T L A L A I R L F G N I M S G S M I V A I L L

 5260 5270 5280 5290 5300 5310 5320 5330 5340 5350 5360 5370 5380 5390 5400
 ACAATTACACCATTCATCTCTGATCTTATGATAGCTCTCGCTCTGGCATGGCAGGGCTACATCTGAGCTTACAGTTCACATTACTGCTGCCACGAGCAGAACATCCAGAAAGTGAAGAACAA
 T I T P F I F L I L *

 (AtpE)
 5410 5420 5430 5440 5450 5460 5470 5480 5490 5500 5510 5520 5530 5540 5550
 ATTGAATAGCAAATCTAACACGGAAAGGCACTAGCTGGCTTGGACACCTATAACTACAATCGGGCTGGCATATTGGCAACTTCTGGCATCAACATAGGCAATTGGGTTATGGCAGGCCAACATTGG
 M A L D T Y I T T I A V A S I A T S G I T I G I G V I G P A I G E G R A V A

 5560 5570 5580 5590 5600 5610 5620 5630 5640 5650 5660 5670 5680 5690 5700
 ACAGCATTTGAGCTCACTGGCACAACACCCGATGCTCCGCAACAATAACCGGACTTTGGGACTGGCTATGATTGAGTCTCTGGCAATCTACTGTTGTTCTGATTCATTTGCAACCCGCTGGACCGT
 T A L S S L A Q Q P D A S A T I T R T L F V G L A M I E S L A I Y C F V V S M I L I F A N P F W N R

 (AtpF)
 5710 5720 5730 5740 5750 5760 5770 5780 5790 5800 5810 5820 5830 5840 5850
 GCACTTACCTGAAATGGAGGAGAAAGTAGACTATGCTGATGTTGCTCAGGTAATCAACTCTCTCATGGCTGTTGGTGAAGCGTCTTACAGGCTATACTCAACGCGAGCAGCGAGAATA
 A L T *
 M L I D W F T V I A Q V I N F L I L V W L L K R F L Y K P I L N A V D A R E N K

 5860 5870 5880 5890 5900 5910 5920 5930 5940 5950 5960 5970 5980 5990 6000
 AGTCGCAAATGAACTTAAGAATGCGGATGCAAAAGAACAGAGCTAAAAGAAAAGAAGATTCAAGCGTAAGAACGAAGAGTCGACCGAGCGCTATGACCTCTGAAACAAAGGAAGAACAGCGAACG
 V A N E L K N A D A K E A E A Q K E K E E F K R K N E E F D Q Q R N D L L N K A K E E A Q A E R Q R

 6010 6020 6030 6040 6050 6060 6070 6080 6090 6100 6110 6120 6130 6140 6150
 GCTTTGAAAGCTGAGGGAGGACTCTCTATTGAGGCAAACACAGGGCATTAAGGAAGCAGAACATTAAAGGAGCAGATTAAAGGAGCAGGAGCTATGAGCTGGCTAGACCCAAAGGAAGTCT
 L F E D V R E E A S Y L R A K Q Q E A L R N E G Q N L S Q E I G R Q T Q Q E V F S I A R K I L T D L

6160 6170 6180 6190 6200 6210 6220 6230 6240 6250 6260 6270 6280 6290 6300
 TGACTGAAACCTGAGAAGAACCGCCAGTGTGTTGTTCAAGGCTGCCAATCTGCAAGATAATGAAAAGACGAGCTTGAGCTCAGCCCTTAGATCATCACAGGTCAGGTGCTACCGTACTGCATTGATCTTCCACAAAT
 T E T S L E E R A V D V F V Q R L R N L Q D N E K K Q L A S A L S T S P G Q V L I R T A F D L P Q T

 6310 6320 6330 6340 6350 6360 6370 6380 6390 6400 6410 6420 6430 6440 6450
 CGCAGCGTGTGTTCTATTAAAAGGGATAAAGAAACTCTGATATCGAAACCTCAACCCAGGTTGAGACTTACAGACCTTGTCACTGGTATTGAATAACAGACGGACAAAAGTTGCGTGGAGCATTCAGATTCAT
 Q R D S I K K A I K E T L D I E T Q P R F E T L P D L V S G I E L N T D G Q K V A W S I A D Y L T S

 6460 6470 6480 6490 6500 6510 6520 6530 6540 6550 6560 6570 6580 6590 6600
 CGATGAAAAAAAGTATGACGAACTGCTGAGCAGCAGCTGAGGAAATCCAACTCTGAGCTCAAACAAAATAAAAGCTAAACCGAACCTCAACCAAATAACAAACCCAGGTTGAAACAGAACAGGCTGAAAGG
 M Q K S I D E L L N E Q P E S K S K S E P Q T K N N K A K T E P Q T K N N K P R F E N E P E A E Q E

 6610 6620 6630 6640 6650 6660 6670 6680 6690 6700 6710 6720 6730 6740 6750
 AGCTTGAACTGAACTAATGACCCAGCCGAAAAGCTGAGCTAGAGTTAACCCAGGCTGCACTGAGTATAATACTCCAGATCAGAAAAGAGCTAGAGACCGAACAGAGTTGATGCAAAACAAGAGCTGAGGCTAGACAAG
 L E A E T N E P R P K T E S E V K T R P A T E Y N N P R S K K E L E T E Q E F D A K Q E S E A R Q E

 (AtpA)
 6760 6770 6780 6790 6800 6810 6820 6830 6840 6850 6860 6870 6880 6890 6900
 AGCTGATGCAAAACAGAGTCAGGGCTAAACAAAATCTGAGCTGAAACAAATGCCGATGCAAAACAGGGCTAGAGCAAGAGTCAGAGGTTAACAGCAATCTGAAATAGTCTTAATCTGCAAGATGAGATAAT
 P D A K Q E S E A K Q K S E A E Q M P D A K Q G S E A K Q E S E V K Q Q S G N S P N L W P R *
 M R L I

 6910 6920 6930 6940 6950 6960 6970 6980 6990 7000 7010 7020 7030 7040 7050
 CCCCCAACAGGCCCCACATAAGAGCCAGTGAACATGGAACTAAAGCTTAAAGACGTTTCGATAAAGTTTACCGGAAATACATCAGGTCAGGGAACTAGTGCACCCGACGCTTACCCGAGGGTAGGCAAGATTGACAGT
 P N Q S P H K E P V N M E T K S L K D V F D K V F T E I H Q V R E S Y T P T L T P K E V G R I L T V

 7060 7070 7080 7090 7100 7110 7120 7130 7140 7150 7160 7170 7180 7190 7200
 TTCTACAGGTATGCCAACTGATCTGGCTTCCACTGTAGGTTTGATGAACTTATAAGTTCCAGGGATTGTTGGATGCTTAATGTTGATGAAAAGAAATAGTACCGTCTACTGGGTGAAATACTCACCTGATGC
 S T G I A N V S G L P T V G F D E L I K F P G D L F G I A F N V D E K E I G T V L L G E Y S H L H A

 7210 7220 7230 7240 7250 7260 7270 7280 7290 7300 7310 7320 7330 7340 7350
 TGGAGATCAGGTTGAAACGCACTAGACGGTTATGGATGCTGGCGTAGGCGAAGAATTGCTGGACGAGTATTGACCCACTAGGTGTCGTCACCTGATAACAAAGGCAATAGTTCCGAACTAACGCTTGCCTATTGAAACGGCCAGTC
 G D Q V E R T R R V M D V A V G E E L L G R V I D P L G R P L D N K G P I V S S S K R L P I E R P S P

 7360 7370 7380 7390 7400 7410 7420 7430 7440 7450 7460 7470 7480 7490 7500
 CGCAATTATGGATCAGGCTCTGTTACCGTCTTCAGACGGCTGCAAAGTTATGTCGCTTAACCAATAGGACCGGCTAGAGAGAAATTGTTAGGGACCGCAGACTGGCAAAACTGCAATTGCAAGTATGATACTCCT
 A I M D R A P V T V P L Q T G V K Y I D A L I P I G R G Q R E L I L G D R Q T G K T A I A I D T I L

 7510 7520 7530 7540 7550 7560 7570 7580 7590 7600 7610 7620 7630 7640 7650
 TAATCAGCGGAAATTAAATGCTCTGGCTTATTGTCGATTCGGCTAGCGTCACTGGCGAGTGGCGAACAGCTGGCAACAACTTACGGGAAAGGGCGCAATGGATTACAGATTGTTGTAACCTGAGGCAATGACCATCAGGGT
 N Q R N F N V L C V Y C A I G Q R A S A V A R A V A N L R E R G A M D Y T I V V V T E G H D P S G L

 7660 7670 7680 7690 7700 7710 7720 7730 7740 7750 7760 7770 7780 7790 7800
 ATCTATATCTCTTATGCTGGACAGCATTGCAAGAGTATTTATGGAATCAGGTGAGACGAGCTGTAATTGTTACGATGACCTGACCAATCATGCACTGTCGCTATCGAGAACTTCCCTTTACTCTGCGCCTCCTGGAGG
 I Y I T P Y A A T S I A E Y F M E S G R D V L I V Y D D L T N H A R A Y E R L S L L L R R P P G R E

 7810 7820 7830 7840 7850 7860 7870 7880 7890 7900 7910 7920 7930 7940 7950
 ACCGTATCCGGTGACATTTATTTACTCAGCACTACTGGAGCGCTACGGCACCTGCTCAAAGAATTGGTGGCGGATCAACTGCTCTCTATTATCGAAACCGAGCGAGAATTTCGCTTATTCGAACAACT
 A Y P G D I F Y I H S R L L E R S T H L L K E F G G G S L T A L P I T E T E A Q N I S A Y I P T N L

 7960 7970 7980 7990 8000 8010 8020 8030 8040 8050 8060 8070 8080 8090 8100
 GATTTCAATTACGGACGGCAGATTACCTTACCTCGCTTGGCTTGGCTGAGCTGGCTGCTGGCGAACATCCGCTCTCGTGTAGGGCGTAAGCAGCTGCTCTGCTATCGAGAAGTGGCTGAGATCTCAAGCT
 I S I T D G Q I Y L S P S L F E L G V L P A V D V G K S V S R V G G K A Q L A A Y R E V A G D L K L

 8110 8120 8130 8140 8150 8160 8170 8180 8190 8200 8210 8220 8230 8240 8250
 TCCCTATTCAAAATTGAAAGCTGAAACGTTACCGGATTCCGGCAAAGCTGAGCTAGAAATACAGGAAGACTCATAGAACATGGAGGGATCGCCCTCTGAGCAGCTCATTAATCTCCGTAACCTGATCAGAT
 A Y S Q F E E L E T F T R F G A K L D E N T R R L I E H G R R I R A L L K Q P H N S P V P V P D Q I

 8260 8270 8280 8290 8300 8310 8320 8330 8340 8350 8360 8370 8380 8390 8400
 GTTATTCATGCTGATTGAAACGCAAACCTCTGATAATGTCGCGCTTGACAAAATGAGCTAGGGCTGAAACTCCCTAGCTGAAGGATTTCTGAAGATGTCGCGAGAGGTTCAAAGGCGATAAAGGAGATAA
 V I L V A L N A K L F D N V P L D K M V E A E N S L R K A L P D I P E D V R E R F K G D K E L S D K

 8410 8420 8430 8440 8450 8460 8470 8480 8490 8500 8510 8520 8530 8540 8550
 GGACCGTGAACATATCCTCAATGCCCGCAAAGCACTGGAACTTACCGCCAAGCTGCAATCCGAACTTAAGCCGAGCTAAACGGAGGAAAACGGCGAGTCAGAAACCAAACGAGATACTGAGTCAGAGCTAAAC
 D R E T I L N I A R K A L E P Y Q P K P E S E S K P E A K T E E N A E S E T Q T E D K S E S E A K T

 8560 8570 8580 8590 8600 8610 8620 8630 8640 8650 8660 8670 8680 8690 8700
 TGAAGAAAAGTGGGAAATCAGAGCTCAGACTAAAGGTAACCTGAGACAGAAAACGAGCTGAGGAAAAGCTGAGTGGAGCCGACGGAAAAGGAGAAAATAAGTAATCAAAGCCAGACTAAAGGCGAGTCAG
 E E K V A K S E T Q T K G K P E T E T K T E E K P E F G T R T E E K *

 (AtpG)
 8710 8720 8730 8740 8750 8760 8770 8780 8790 8800 8810 8820 8830 8840 8850
 AAGGGAAAATCATGGCGAACACTCAGAGTTGCGACAAAGGATCTGAGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAG
 M T E T T Q S L R T K I D R A K D L Q S V V R T M K A L A A S N I G Q Y E K S V S A L S D Y Y

 8860 8870 8880 8890 8900 8910 8920 8930 8940 8950 8960 8970 8980 8990 9000
 ATTATACTGGAACTGGGTTGGCTTATGTTCAAGGAAATTGCACTGCAAGCTCCACAGGAAGGAAAGCGCAAAGAAATACAGCTTACGCTTACGGAGCTGCTGTTGGTTCGGTACGATCAGGGCTTGTGGGCTAG
 Y T V E L G L G L C F R K I A F M P A P Q E G K A Q K N T R L I G A V V F G S D Q G L V G Q F N D I

 (AtpJ)
 9010 9020 9030 9040 9050 9060 9070 9080 9090 9100 9110 9120 9130 9140 9150
 TAATTACCGATTATGCGGCTAGGAACTGAAAGCTTGGCAGGAAAGCTGAGGTTGGCTGAGGAGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAG
 I T D Y A V R E L K A L P G K A Q V W A V G E L V I H A W K M K A C R L L D F I M C Q I R L R Q L P
 M G C R G A C Y S R L E D E G L P L I G L Y N V P N S V K A I T

 9160 9170 9180 9190 9200 9210 9220 9230 9240 9250 9260 9270 9280 9290 9300
 CCCGTTATAGCGCAAATTCTGAGGAAATACCGCAGCCAGGATGAAGATGCCAACTTCATCTTACCAACCCCAACAAACGAGTTACCTAGGCGAGTCAGCCAGGACTGCTGAGGCTTGTGGGCTAG
 R L *
 P L I A Q I L V E N E K L R S Q D E D A E L H L Y Y N R H K T R V T Y E P V S Q R L L P F D E T W R

9310 9320 9330 9340 9350 9360 9370 9380 9390 9400 9410 9420 9430 9440 9450
AGACGCCCTGACTAACTTCTGGCCAACCAAAAACCTTGCTGCGTGAGGTAATGGCAATATCACAGAACCTTACGGGACTCATCCGGAGTACCTTCGTTCGCTTCAAGGGCATGCCGCTGAGTCCCTGCAAGTGAAAACTC
D D L T K L S W P T K N L L P E V M G N I T E T L R A L I R E Y L F V S L F R A C A E S L A S E N S

9460 9470 9480 9490 9500 9510 9520 9530 9540 9550 9560 9570 9580 9590 9600
TAGTCGACTGGCAGCAATGCAGCGTGGCGATAAAAACATTGGAGGAGCTGCTTAGTAACTCTAGTGGAGAATACCCGTGCGCCAGAGCGGCATTGATGAGGAGTTGTTGAAGTTGTTCAAGGTTGCTGAGCTCTGTCAGGGCCCG
S R L A A M Q R A D K N I E E L L S N L S G E Y Y R V R Q S G I D E E L F E V V S G F E A L S R A R

9610 9620 9630 9640 9650 9660 9670 9680 9690 9700 9710 9720 9730 9740 9750
TGTGAAATAAGAAAAACTTGTGAAATAAGGGAGCTTGTGAAATAAGGAAAGCCGTTATTTAAGAGAACCTTGTGACATTTCAGGAAATAATTTGTAATAAAAGAGTTTGTGAAATAAGAGACTTGTGAAATAAGAGACTTGTGAAATAAGAAAAACT
S S R R *

9760 9770 9780 9790 9800 9810 9820 9830 9840 9850 9860 9870 9880 9890 9900
TGTGAAATAAGAAAAACTTGTGAAATAAGGGAGCTTGTGAAATAAGGAAAGCCGTTATTTAAGAGAACCTTGTGACATTTCAGGAAATAATTTGTAATAAAAGAGTTTGTGAAATAAGAGACTTGTGAAATAAGAGACTTGTGAAATAAGAAAAAT

9910 9920 9930 9940 9950 9960 9970 9980 9990 10000 10010 10020 10030 10040 10050
CTAGGAGCTGCTGGAAAATTCAAATTAACTATTAGATAATTATTCTTAAAAAATTATATTATGAGTTGGTACCGATAGTGCAGTAGACAGAGCTCTCAAGAACAGAACCTTCTGAGCTTTGTTGA

10060 10070 10080 10090 10100 10110 10120 10130 10140 10150 10160 10170 10180 10190 10200
CTGGAAATGGTAAACTCGCAATCATTGTTTCTGTTAGGGGGTTTCGAGTTATGGAAATTCGAGCGATAACAATAGCTTCAAGAACCTTCTGAGTAACTTCTGAGTAACTTCTGAGTAACTTCTGAGTAACTTCTGAGTAA

10210 10220 10230 10240 10250 10260 10270 10280 10290 10300 10310 10320 10330 10340 10350
CATGCATAATATTGCTTTCAGTTATTTATCAGGTGCCAACCTGGCAACTGTAACCGCAATAATAGCTTCCCTTATTCGCTTACATTTCAGTATTATGGAGTTGTTGTGGAATCCCCTTGTACGAA

10360 10370 10380 10390 10400 10410 10420 10430 10440 10450 10460 10470 10480 10490 10500
CGAAGTAAAATTCTGGCTTATTCAGGGAGTTCTGGAAAAGGATCCGCTCTTGCTTGCGTTAGCGATTGGTCTGGTACTTCGCGCTTCGGAATATCCCTTCTCTTACCCCTTATTCGTGAGGATCCCTGAGA

10510 10520
CTTTCTTGTCTGCCAGTCGGAGG

付録B *M. barkeri* F タイプ ATPase 遺伝子産物と既知の F タイプ構成構成成分との相同性

1st Amino Acid Sequence

File Name : atpA.aa
Sequence Size : 588

2nd Amino Acid Sequence

File Name : atpA Moorella thermoacetica
Sequence Size : 507

Unit Size to compare = 1

[51.3% / 86.6% / 464 aa]

1'	MRLIPNQSPHKEPVNMETKSLKDVFDKVFTIEHQVRESVTPLTPKEVGRILTGSTGIAN
1"	MSIRPDEITSILKNQIEQYQLEVEMAEVGTVTQVGDGIAR
61'	VSGLPTVGFDELIKFPGDLFGIAFNVDKEIGTVLLGEYSHLHAGDQVERTRRVMDVAVG
41"	IYGLDRAMAGELLEFPGDIYGMVLNLEEDNVGAVILGPYTHIKEGDQVKRTGRIVEVPVG
121'	EELLGRVIDPLGRPLDNKGPIVSSKRLPIERPSAIMDRAPVTPLQTGVKVIDALIPIG
101"	EALIGRVVNAMQPIDGKGPIQTDKFRPVEPAGVYRQPVNTPLQTGLKAIDSMVPIG
181'	RGQRELIILGDRQTGKTAIAIDTILNQRNFNVLCVYCAIGQRASAVARAVANLRERGAMDY
161"	RGQRELIIGDRQTGKTAIAVDTIINQKGQNVICIVYAIGQKASTVAGVVQRLEEAGAMEY
241'	TIVVVTEGNPDSGLIYITPYAATSDIAEYFM-ESGRDVLIVYDDLTNHARAYRELSLLLRR
221"	IIVVMATASEPAPMLIYAPIYAGCTMGEYFMYEQHRDVLCVYDOLSKHAAAYRELSLLLRR
300'	PPGREAYPGDIFYIHSLRLLERSTHLLKEFGGSLTALPIIETEAQNISAYIPTNLISITD
281"	PPGREPYPGDFVYLSGLLERPARLTDSLGGSLTALPVIETQAGDVSAYIPTNVISITD
360'	GQIYLSPSLFELGVLPADVGVKSRSVGGKAQLAAAYREVAGDLKLAYSQFEELTFTRFG
341"	GQIFLESDFYAGQRPAINVAVSVSRVGGAAQIKAMKQVAARLRLDQAQYRELAFAQFG
420'	AKLDENTRRLIEHGRRIALLKQPHNSPVPVPDQIVILVALNAKLFDNVPLDKMVEAENS
401"	SDLDKATQARLARAERMMEILKQDQYQMPVVEQVVVLYAAVNGFLDDLPVGRVRAFEKD
480'	LRKALPDIPEDVRERFKGDKELSKDRETIILNIARKALEPYQPKPESESKEAKTEENAE
461"	FLRFLRNERPEVLAGIREKRQLDDNLQEQLKKSIEDFKGSFTAAGES

1st Amino Acid Sequence

File Name : atpB.aa
Sequence Size : 190

2nd Amino Acid Sequence

File Name : atpB Moorella thermoacetica
Sequence Size : 246

Unit Size to compare = 1

[36.0% / 79.1% / 172 aa]

1'	MRLSPDELIFWQYDFIKLNATIVYTWGLMLVMV * . . * * *
1"	MTGENDANKTRANSLATINMGLRALGEIMTHVRPVEIFHLGPIPIYSTVVNTWIIMILL
34'	IGSKTITSKLSTGLERSRWQNILEIIVTGILEQIEDVGLDQPRKYLGFGLGTLFLFIAVAN * . * *** . * . * . * * . *** . . * . * . * . * . * . * . * .
61"	AGIFLATRKLSF-IPRGA-QHVLEMFLEFFYGLLEIIGKEGRRYLPLVATLFIFILSN
94'	LCIIIPGYEPPTGSLSTTAALALCVFAVPLFGIEEQGISNYLKTYTEPTIIMLPFNIIS * . . *** . *** .
119"	LSWFIPGMKPPTMDLSTTAFAVTTIILVQIFGIRKLGLRGYIRHFFQPAPFLFPLNVIE
154'	EISRTLALAIRLFGNIMSGSMIVAILLTITPFIFLIL * * . *** .
179"	ELVKPVSLSLRLFGNLGEEMVVTILFLMIPFLLPTPIMLLGVLMGTIQAFVFTLLTITY

1st Amino Acid Sequence
File Name : atpC.aa
Sequence Size : 144

2nd Amino Acid Sequence
File Name : atpC Haemophilus influenzae
Sequence Size : 151

Unit Size to compare = 1

[26.9% / 67.7% / 130 aa]

```
1' MNSGLMNSGLMNSGLMNL TILLPFQVFAEKKGVS RIVAEGREGSFGLLPHRLDCVATLGP
      ****. . * . . *..* ..* ***..** . . . .
1" PSWISSPRTMATFNL TIVSAEQKIF-EGEVKQIQVTGVEGELGILPGHTPLTAIKP
61' GILTYETESEGEVYVAVDEGVLIKNGQYVLVSVRDAIFGTDL--SQLHEAVEK--EFLTL
      **.... . . * . * . * . * . * . * . * . * . * . * . * . .
57" GIVKFTLKDGNEEVYVSGGFLEVQPNIVTVLADIAIRGSELDADRIHEAKRKAENIVS
117' DETEQKIRSVMVKLETGLIRR LAEFQNV
      * . . . . * . * . .
117" RGSDADHDLLVAKLSKELA-KLRAYELTEKLLKTRR
```

1st Amino Acid Sequence

File Name : atpD.aa
Sequence Size : 469

2nd Amino Acid Sequence

File Name : atpD C. reinhardtii chl
Sequence Size : 491

Unit Size to compare = 1

[51.8% / 84.6% / 479 aa]

1' MVDRSRLLNLKVISVRGSIVDVLFEK-HLPPVY-TL-LRA---GRESQ
* * . *.*. . *.*. . *.*. . *.*. . *.*. . *.*. . *.*.
1" MPWGILIPLTMSDSIETKNMGRIVQIIGPVLDIVFAKGQVPNIYNALTIRAKNSAGTEMA
44' IAIEVLTQLDAHHVRGIALTPTEGLARGMAVEDTGGPLKAPVGREILSRMFDFVFGNTIDR
. ** * . *.. * . . * . . * . . * . . * . . * . . * . . * .
61" VTCEVQQLLGDNCVRAVSMNPTEGLMRGMEVVDTGKPLSVPVGKVTLGRIFNVLGEPVDN
104' RKPPSDIQWRSIHQTPPPLMRRSTTSEIFETGIKAIDVLVPLERGGKAGLFGGAGVGKTV
. . . . * * * * * *
121" MGNVKVEETLPIHRTAPAFVLDTRLSIFETGIKVVDLLAPYRRGGKIGLFGGAGVGKTV
164' LLTEMIHNVVKQHQGVSIFCGIGERCREGEELYRDMKDAG----VLPNT--VMVFGQMN
. . . . * * * * * *
181" LIMELINNIAKAHGGVSFAGVGERTREGNDLYTEMKESGVIVEKNLSDSKVALVYQMN
217' EPPGARFRVGHTALTMAEYFRDDERRDVLLIONIFRFIQAGSEVGLMGQMPSRLGYQP
*****.**. *****.*****.
241" EPPGARMRVALTALTMAEYFRDVNKQDVLFFIDNIFRFVQAGAEVSALLGRMPSAVGYQP
277' TLGTELSELEERISTTDAGAIMSIQAVYVPADDFTPSAVHTFSHLASIVLSRKASEG
..
301" TLATEMGLQERITSTKDGSITSIQAVYVPADDLTDPAATTFAHLDATTVLSRNLAAG
337' LYPAIDPLQNSKMATPGIVGERHYLLAQEIROTLAQYSELKDIISMLGLEQLSPEDRNV
.
361" IYPAVDPLESTSTMLQPWILGEKHYDSAQSVKTLQRYKELQDIIAILGLDESEEDRLI
397' VARARRLERFLTQPFFTTEQFTGIKGKSISLSDALGCCERILNDEFKDYSEGDLYMI GTI
*****.*****.*****.
421" VARARKIERFLSQPFFVAEVFTGSPGKYVSLAETIEGFGKIFAGELDDLPEQAFYLVGNI
457' DEAMAKKSSREKS
.
481" TEAISKAASLK

1st Amino Acid Sequence
File Name : atpE.aa
Sequence Size : 91

2nd Amino Acid Sequence
File Name : atpE O. neapolitana plastid
Sequence Size : 82

Unit Size to compare = 1

[48.7% / 90.8% / 76 aa]

1'	MALDTYITTIAVASIATSGITIGIGVIGPAIGEGRAVATALSSLAQQPDASATITRTL FV .. * * . . * * . . * * . * . * * . * . * . * . * . * . * . * . * . * . * . * .
1"	MNPIVSGASVVAAGLAAIGPGIGQGTAAAQAVEGLARQPEAEKGIRGTLLL
61'	GLAMIESLAIYCFVVSMILIFANPFWNRALT . * * . * * . * * . * . * . * . * . * . * . * . * . * . * . * . * . * . * .
56"	SLAFMESLTIYGLVVALCLLFANPFAG

1st Amino Acid Sequence
File Name : atpF.aa
Sequence Size : 386

2nd Amino Acid Sequence
File Name : atpF Enterococcus hirae

Sequence Size : 174

Unit Size to compare = 1

[26.0% / 76.0% /150 aa]

1'	MLIDWFTVIAQVINFLILVWLLKRFLYKPILNADV DARENKVANELKNADA*****..***.* . ** . . . **.*.**.*...*
1"	MLNQLAIAEVGNPMLGNIIVVSGSFLILMFLLKHFAWGPISDILKKREDKIANDLDSAEK
51'	KEAEAQKEKEEFKRKNEEFDQQQRNDLLNKAKEEAQAERQRLFEDVREEASYLRAKQQEAL* .. * * .. ***... . **.**... *...*...*
61"	SRINSAKMEQEREQQLLASRSDAADI IKNAKESGELSRQNILKETQEEVARLKSKAQTDI
111'	RNEGQNLSQEIGRQTQQEVFSIARKILTDLTETSLEERAVDVFVQRLRNLQDNEKKQLAS* .. *
121"	MLERDTALNSVKDDVADLSLQIAAKILNKELSPEMHESLINQYIEGLGSSNETR

1st Amino Acid Sequence
File Name : atpG.aa
Sequence Size : 149

2nd Amino Acid Sequence
File Name : atpG Arabidopsis thaliana chl
Sequence Size : 373

Unit Size to compare = 1

[23.4% / 64.0% / 111 aa]

1'	MTETTQSLRTKI
1"***..*
13'	MACSNLTTMWVSSKPSLSADSSSLSFRSVLKCPNTSSPPSRASSVSPLQASLRELRDRI
**...*...**...**...*....*
61"	DSVKNTQKITEAMKLVAAAKVRRAQEAVVNGRPFSETLVEVLYNINEQLQTDDVDVPLTK
73'	QKNTRLIGAVVFGSDQGLVGQFNDIITDYAVRELKALPGKAQVWAVGELVIHAWKMACR
***...*...**...*...*....*
121"	VRPVKKVALVVVTGDRGLCGGFNNFIKKAEARIKELKGLGLEYTVISVGKKGNSYFLRR
133'	LLDFIMCQIRLRQLPRL
181"	PYIPVDKYLEAGTLPTAKEAQAVADDVFSLFISEEVDKVELLYTKFVSLVKSEPVIAHLL

1st Amino Acid Sequence
File Name : atpI.aa
Sequence Size : 112

2nd Amino Acid Sequence
File Name : atpI Rhodospirillum rubrum
Sequence Size : 123

Unit Size to compare = 1

[24.2% / 64.2% / 95 aa]

1' MSDKPSKKPLED-ESPLARQVGTKAERKIRAQRHVDR_TIWLGLGMMGLIGWSVAVPTLIG
.. . *** ... * . . * ***. *
1" MTDRDTPPSLEDISRRLTEAKGGADGAEADGAGSSGPARASGLGIGMRISIELVT_TIAVG
60' AAIGLWLDKHYPEFSWALTMLIIGLIIGCLNAWHWLAKERQEMQEEQEDYNE
*** **. . * . . * . . * . . * . . ***.
61" GAIGYGLDSWLGTSP_LAMVVFLVLGGAAGVMNAYRVVKGLDDSVGLGRAIERKEKAEGNK

1st Amino Acid Sequence
File Name : atpJ.aa
Sequence Size : 186

2nd Amino Acid Sequence
File Name : atpJ Bacillus subtilis
Sequence Size : 290

Unit Size to compare = 1

[31.2% / 70.8% / 144 aa]

1'	MGC
61"	RVSGNVKHPMLLSREVKKTAYLVITSRGLAGAFNSSLRSAYQAMQERHQSKDEYAVIA
4'	RGACYSRLEDEGLPLIGLYNVPNSVKAITPLIAQILVENEKLRSQDEDALHLYYNRHKT
121"	IGRVGRDFFKKREIPIISETGLGDEVTFTIEIKDLARQTIQMFIDGAFDELHLVYNHFVS
64'	RVTYEPVSQRLLPFDETWRDDLTLSWPTKNLLPEVMGNITETLRALIREYLFVSLFRAC
181"	AITQEVTTEKKLLPL----SDLGS GG-GKRTASYEFEPSEEEVLEVLLPQYAESLIFGAL
124'	AESLASENSSRLAAMQRADKNIELLSNLSGEYYRVRQSGIDEELFEVVSGFEALSRARS
235"	LDSKASEHAARMTAMKNATDNAKELIDSLSYNRARQAAITQEITEIVGGAAALE

1st Amino Acid Sequence
File Name : atpJ.aa
Sequence Size : 186

2nd Amino Acid Sequence
File Name : atpJ Bacillus subtilis
Sequence Size : 290

Unit Size to compare = 1

[31.2% / 70.8% / 144 aa]

1'	MGC
61"	RVSGNVKHPMLLSREVKKTAYLVITSRGLAGAFNSSLRSAYQAMQERHQSKDEYAVIA
4'	RGACYSRLEDEGLPLIGLYNVPNSVKAITPLIAQILVENEKLRSQDEDAELHLYYNRHKT
121"	IGRVGRDFFKKREIPIISETGLGDEVTFTEIKDLARQTIQMFIQGAFDELHLVYNHFVS
64'	RVTYEPVSQLLPFDETWRDDLTCLSPTKNLLPEVMGNITETLRALIREYLFVSLFRAC
181"	AITQEVTTEKKLLPL----SDLGS GG-GKRTASYEFEPSEEEVLEVLLPQYAESLIFGAL
124'	AESLASENSSRLAAMQRADKNIEELLSNLSGEYYVRQSGIDEELFEVVSGFEALSRARS
235"	LDSKASEHAARMTAMKNATDNAKELIDSLSLSYNRARQAAITQEITEIVGGAAALE

付録C VおよびFタイプ、触媒および非触媒サブユニットのアライメント

vvr_Ncr	+								-	MVLSDKE
vvr_Sce	-	-	-	-	-	-	-	-	-	MALRAMRGIVNGAAPELPVPTSGLA
vvr_Bta	-	-	-	-	-	-	-	-	-	MNAQQ
vvr_Dme	-	-	-	-	-	-	-	-	-	MAP
vvr_Hvu	-	-	-	-	-	-	-	-	-	MLSKQE
vvr_Tco	-	-	-	-	-	-	-	-	-	-
avr_Mba	-	-	-	-	-	-	-	-	-	-
avr_Mma	-	-	-	-	-	-	-	-	-	-
avr_Hsa	-	-	-	-	-	-	-	-	-	-
avr_Hvo	-	-	-	-	-	-	-	-	-	-
avr_Mja	-	-	-	-	-	-	-	-	-	-
evr_Ehi	-	-	-	-	-	-	-	-	-	-
evr_Tth	-	-	-	-	-	-	-	-	-	-
avr_Sac	-	-	-	-	-	-	-	-	-	-
cfr_Osa	-	-	-	-	-	-	-	-	-	MATLRVDEIHKI-----LR
cfr_Zma	-	-	-	-	-	-	-	-	-	MATLRVDEINKI-----LR
cfr_Tae	-	-	-	-	-	-	-	-	-	MATLRVDEIHKI-----VR
cfr_Nta	-	-	-	-	-	-	-	-	-	MVTIRADEISNI-----IR
cfr_Sol	-	-	-	-	-	-	-	-	-	MATIRADEISKI-----IR
cfr_Psa	-	-	-	-	-	-	-	-	-	MVTIRADEISQI-----IR
cfr_Mpo	-	-	-	-	-	-	-	-	-	MVNIRPDEISSI-----IR
cfr_Cpa	-	-	-	-	-	-	-	-	-	MVSIRPDEISSI-----IR
cfr_Ppu	-	-	-	-	-	-	-	-	-	MVNIRPDEISSI-----IR
efr_Syn	-	-	-	-	-	-	-	-	-	MYSIRPDEISSI-----IR
efr_Bsu	-	-	-	-	-	-	-	-	-	MSIKAEIISTL-----IK
efr_Ehi	-	-	-	-	-	-	-	-	-	MAIKAEIISAL-----IK
efr_Pmo	-	-	-	-	-	-	-	-	-	-
mfr_Osa	-	-	-	-	-	-	-	-	-	MEFSPRAEELTTL-----LE
mfr_Zma	-	-	-	-	-	-	-	-	-	MEFSPRAEELTTL-----LE
mfr_Bta	-	-	-	-	-	-	-	-	-	-
mfr_Dme	-	-	-	-	-	-	-	-	-	MLS VRYAAA VVAR ALPRAGLV-SKN ALGSS FIA ARNL HAS NSRL QKT GTAEVSSI-----LE
mfr_Hsa	-	-	-	-	-	-	-	-	-	MLS VRYAAA VVAR ALPRAGLV-SRN ALGSS FIA ARNL HAS NSRL QKT GTAEVSSI-----LE
mfr_Spu	-	-	-	-	-	-	-	-	-	MSIF SRLASS VARN LPKA ANQVACK AAYPA ASLA ARKL HVAST Q-----RSAE ISNI-----LE
mfr_Ncr	-	-	-	-	-	-	-	-	-	MAQ VSV RLAAA AARSL SRQAP QVA-KRCIGAAYVSHRN VTTS APR---PSTA EVSSI-----LE
efr_Rbl	-	-	-	-	-	-	-	-	-	MFR NAL RQSTR AVG AF SAT GRVA ARNA APV VSA VQARTY ADA-----KAT PT E VSSI-----LE
efr_Cre	-	-	-	-	-	-	-	-	-	MG IQA-----AE ISAI-----LK
efr_Sli	-	-	-	-	-	-	-	-	-	MRS QAL SLARAG LLQL SSQT GAS LEGG FALS KRAE QAL I RAS RAFA SDAK AL DEL RKP KFT SKYL INH VS-----MAEL TIR PEE I RD-----ALE
efr_Eco	-	-	-	-	-	-	-	-	-	-
afr_Mba	-	-	-	-	-	-	-	-	-	MQL NSTE ISEL-----IK
cfc_Osa	-	-	-	-	-	-	-	-	-	MRL IPN QSPH KEP VN MET KSL KD VFD KV F-----
cfc_Zma	-	-	-	-	-	-	-	-	-	-
cfc_Tae	-	-	-	-	-	-	-	-	-	-
cfc_Nta	-	-	-	-	-	-	-	-	-	-
cfc_Sol	-	-	-	-	-	-	-	-	-	-
cfc_Psa	-	-	-	-	-	-	-	-	-	-
cfc_Mpo	-	-	-	-	-	-	-	-	-	-
cfc_Pth	-	-	-	-	-	-	-	-	-	-
cfc_Cpa	-	-	-	-	-	-	-	-	-	-
cfc_Ppu	-	-	-	-	-	-	-	-	-	-
efc_Syn	-	-	-	-	-	-	-	-	-	-
mfc_Osa	-	-	-	-	-	-	-	-	-	MAT RRAL LSS LVRAA-----SRL RGASP-APR PRGPL HRP SP SG YLF FN RAA AYATA AAA AKER P--PAP
mfc_Zma	-	-	-	-	-	-	-	-	-	MAS RRV YSS LLS RSA-----SRL RAAS PAAP PRP RAPP HRP SP AG YLF FN RAA AYASS AAAA QAA PAT TPP
mfc_Nta	-	-	-	-	-	-	-	-	-	MAS RRL L ASLL RQSA QR GGG L IS RSL GNS IPKS A RASS-RAS PKG FLL N RAV QY AT SAA APAS QP ST PP
mfc_Hbr	-	-	-	-	-	-	-	-	-	MAS RRL L SLL R SSS R RSV KSP ISN INPK L SSS PSS KSR AS P YG LL TRA EY AT SAAA APP QPPP A
mfc_Bta	-	-	-	-	-	-	-	-	-	ML GL VGR VY AAS AS GAL RGL SP A PL PQA Q L L R A A P AA
mfc_Rno	-	-	-	-	-	-	-	-	-	-
mfc_Sce	-	-	-	-	-	-	-	-	-	-
efc_Rbl	-	-	-	-	-	-	-	-	-	MVL PRL Y TATS RAA FKA K QSA PLL ST SW KRC MASA A
mfc_Ncr	-	-	-	-	-	-	-	-	-	-
efc_Eco	-	-	-	-	-	-	-	-	-	-
efc_Pmo	-	-	-	-	-	-	-	-	-	-
efc_Bsu	-	-	-	-	-	-	-	-	-	-
efc_Ehi	-	-	-	-	-	-	-	-	-	-
efc_Sli	-	-	-	-	-	-	-	-	-	-
afc_Mba	-	-	-	-	-	-	-	-	-	-
VVC_Ncr	-	-	-	-	-	-	-	-	-	MAP Q
VVC_Sce	-	-	-	-	-	-	-	-	-	MAGA
VVC_Hvu	-	-	-	-	-	-	-	-	-	-
VVC_Zma	-	-	-	-	-	-	-	-	-	-
VVC_Bta	-	-	-	-	-	-	-	-	-	-
VVC_Dme	-	-	-	-	-	-	-	-	-	-
VVC_Tco	-	-	-	-	-	-	-	-	-	-
avc_Mja	-	-	-	-	-	-	-	-	-	-
evc_Ehi	-	-	-	-	-	-	-	-	-	-
avc_Mba	-	-	-	-	-	-	-	-	-	-
avc_Mma	-	-	-	-	-	-	-	-	-	-
avc_Hsa	-	-	-	-	-	-	-	-	-	-
avc_Hvo	-	-	-	-	-	-	-	-	-	-
evc_Tth	-	-	-	-	-	-	-	-	-	-
avc_Sac	-	-	-	-	-	-	-	-	-	-

+ + + + +
 --- MADPRVPSSYNTPIRINYNTVGGVNGPLVILDNVKFPRYNEIVTL--- TLPDGTVRSQGVVL
 vvr_Ncr LFAINKKAVEQGFNVKPRLNNTYNSVGNGLVLILEKVKFPRYNEIVNL--- TLPDGTVRSQGVVL
 vvr_Sce GSREQALAVSRNLYSQPRLTYKTSGVGNGLVLIDHVKFPRYAEIVHL--- TLPDGTRKRSQGVVL
 vvr_Bta AQRHVLAVSRDFISQPRLTYKTSGVGNGLVILDEVKFPKFAEIVQL--- RLADGTVRSGQVL
 vvr_Dme EMEEGTLEIG----- MEYRTVSGVAGPLVLVIDKVKGPKYQEIVNI--- RLGDTTRRRGQVL
 vvr_Hvu LLTEHINELRKDYYVKPRLEYTTIRAVNGPLVLGDVRLPTFAEIVNI--- ELSDGTMRRGQVL
 vvr_Tco ----- MVKEYKTQIAGPLLVFAEVDEPIGYDEIVEI--- NLPDGTTTREGEV
 avr_Mba ----- MVKEYKTITQIAGPLLIFVEKTEPVGYKEIVTI--- KMGDGTVRRGQVL
 avr_Mma ----- MKEYQTITEVSGPLLVYVETDEPIGYDEIVNI--- ETPNGDKVRKGQVL
 avr_Hsa ----- MKEYQTITEISGPLLVFAEVDEPIGYDEIVEI--- ICPDGKRMGQVL
 avr_Mja ----- MATAASAIYE55VKSIAGPLLIVEGVEGAAYGEIVEV---
 evr_Ehi ----- MIKEYRTIKEVVGPLMAVEKVSGVKYEEIVLIEV--- RMQNGEIRRQVL
 epr_Tth ----- MDLLKKEYTGITYISGPLLIVENAKDLAYGATVDI--- KDGTRVRRGQVL
 avr_Sac ----- MSLLNVREYSNISMIGPLIAVQGSDAAYNELVEI--- EMPDGSKRRGLV
 cfr_Osa ERIEQYNNRKVG----- IENIGRNVQVGDDGIARIIGLGEIMSGELVEF--- AEGTRGIAL
 cfr_Zma ERIEQYNNRKVG----- IENIGRNVQVGDDGIARIIGLGEIMSGELVEF--- AEGTRGIAL
 cfr_Tae ELIEQYNNRKVG----- IENIGRNVQVGDDGIARIIGLGEIMSGELVEF--- AEGTRGIAL
 cfr_Nta ERIEQYNNREVK----- IVNTGTVLQVGDDGIARIHGLDEVMAGELVEF--- EEGTIGIAL
 cfr_Sol ERIEGYNREVK----- VVNTGTVLQVGDDGIARIHGLDEVMAGELVEF--- EEGTIGIAL
 cfr_Psa KRIEQYNTTEVK----- IVNTGTVLQVGDDGIARIYGLDDVMAGELVEF--- KETGTVGIAL
 cfr_Mpo KQIEQYNNQEVK----- IVNIGHTVLOVGDDGIARIYGLDKVMAGELVEF--- EDGTVGIAL
 cfr_Cpa QQIEQYDQEIQ----- VSNVGTVLQVGDDGIARVYGLDKVMSELLEF--- EDGTIGIAL
 cfr_Ppu QKIEKYDQDV----- VANIGTVLQVGDDGIARVYGLDEVMAGELVEF--- EDKTVGIAL
 efr_Syn QKIEQYEQSIN----- VDNVGTVLQVGDDGIARVYGLDKVMASELVEF--- EDGTVGIAL
 efr_Bsu QQIQNYQSDIE----- VQDVGTVIQVGDDGIARVHGLDNCMAGELVEF--- SNGVLGMAQ
 efr_Ehi EKIENYQNVL----- VEEIGTVTYVVGDDGIARAHGLENAMELLEF--- SNGSYGMAQ
 efr_Pmo
 mfr_Osa SRMTNFYTTFQ----- VDEIGRNVSVVGDDGIARVYGLNEIQAGEMVEF--- ASGVKGIAL
 mfr_Zma SRMINFYTNLK----- VDEIGRNVSVVGDDGIARVYGLNEIQAGEMVEF--- ASGVKGIAL
 mfr_Bta ERILGADTSVD----- LEETGRVLSIGDGIARVHGLRNVQAEEMVEF--- SSGLKGMSL
 mfr_Hsa ERILGADTSVD----- LEETGRVLSIGDGIARVHGLRNVQAEEMVEF--- SSGLKGMSL
 mfr_Dme ERILGVAPKAD----- LEETGRVLSIGDGIARVYGLNNIQADEMVEF--- SSGLKGMSL
 mfr_Spu ERILGHPTKEN----- LEETGRVLSIGDGIARVYGLKNIQAEEMVEF--- SSGLKGMSL
 mfr_Ncr QRIRGVQEESN----- LAETGRVLSVGDDGIARVHGMANVQAEELVEF--- ASGVKGMC
 efr_Rbl EQIKNFGKDAE----- VAEVGRVLSVGDDGIARVYGLDNVQAGEMVEF--- PGGIRGMAL
 mfr_Cre EKLIPAVKEWEKQYQPPVIIHLGRVLSVGDDGIARVYGLKSQAGELVCF--- DSGVKGMAL
 efr_Sli NFVQSYKPDAASR----- EEVGTVTLAGDGIAKVGEPLPSAHANELLKF--- EDGTLGLA
 efr_Eco QRIAQFNVVSE----- AHNEGTTIVSVDGVIKRVIRIHLGADCQMGEMLSL--- PGNRYAIAL
 afr_Mba TEIHQVRESVTPTLT----- KEVGRILTVSTGIVNGLSPPTVGFDELIK--- PGDFGIAF
 cfc_Osa MRTNPPTS SRPGVSTIEEKSTGRIDQIIGPVLDVTFP-RGKLPYIYNALVVKSROTDGKQ--- INVTCEVQ
 cfc_Zma MRTNPPTS SRPGISTIEEKSVGRIDQIIGPVLDITFP-PGKLPYIYNALIVKSROTDKQ--- INVTCEVQ
 cfc_Tae MRTNPPTS SPVGASTIEEKSTGRIDCIIIGPVLDVTFP-PGKLPYIYNALVVQSRDTDDKQ--- INVTCEVQ
 cfc_Nta MRINPPTS SGGVYSTLEKKNPGRVQVCIIGPVLDVAFP-PGKMPNIIYNALVVQGRDVSQGP--- INVTCEVQ
 cfc_Sol MRINPPTS SDPGVSTLEKKNLGRIAQIIGPVLDVAFP-PGKMPNIIYNALIVKGDRTAGQP--- MNVTCEVQ
 cfc_Psa MTITPPPSDTEVSYLENKNLGRITQIIGPVLDVVFPP-PGKMPYIYNALIVQGRDTVGKQ--- INVTCEVQ
 cfc_Mpo MKTNFLAF-- GMSTLVAKNIGSITQVIGPVLDVAFS-PGKMPNIIYNSLIVKDQNSAGEE--- INVTCEVQ
 cfc_Pth MRKNPLVL-- GVASASVETNVGRIAQIIGPVLDVSPF-PGNMPNIIYNSLIVKGQGTAGQE--- IQVTCEVQ
 cfc_Cpa ----- MATTTSKTNTGYTVQIGPVLDVSPF-NGQLPKIYNAITVKGKNEAGQD--- ITVTCEVQ
 cfc_Ppu ----- MVSTTKS--- TGSVYQTIQIGPVLDIAFP-NGQLPKVFNALIVQGST--- D-GTITCEVQ
 efc_Syn ----- MVTTAERTNVGFITQVIGPVVIDIEFP-SGKMPAIYNALRIQGKNAAGLD--- VAVTCEVQ
 mfc_Osa ATGKATGGGKITDEFTGAGAIQVQCQVIGAVVVDVRFD-EG-LPPILTAEVL--- DHN-IRLVLEVA
 mfc_Zma ATGK-TGGGKITDEFTGAGAIQVQCQVIGAVVVDVRFD-EG-LPPILTAEVL--- DNN-IRLVLEVA
 mfc_Nta KSG-SEPSGKIDTEFTGAGSIGKVCQVIGAVVVDVRFD-EG-LPPILTAEVL--- DNQ-IRLVLEVA
 mfc_Hbr KPEGGKGGGKITDEFTGKGAIQVQCQVIGAVVVDVRFD-EG-LPPILTSELEV--- DHS-IRLVLEVA
 mfc_Bta LQPARDYAAQASPSPKAGATTGRIVAVIGAVVVDVQFD-EG-LPPILNALEVQ--- GRE-TRLVLEVA
 mfc_Rno ----- EFRTATGQIVAVIGAVVVDVQFD-EG-LPPILNALEVQ--- GRE-SRLVLEVA
 mfc_Sce ----- QSTPITGKVTAIVGIAIVDVHFE-QSELPAILNALEIK--- TPQ-GKLVLEVA
 mfc_Rbl ----- MTMATTVSKGKVTVQVIGAVVVDVQFD-EGLVPAILNALETT--- NNG-KKLLLEVA
 mfc_Ncr F----- ASSAGVGDGKIVQVIGAVVVDVQFD-TDKLPPILNALEVQ--- NNG-QKLLLEVA
 efc_Eco ----- MATGKIVQVIGAVVVDVQFD-QDAVPRVYDALEVQ--- NGN-ERLVLEVA
 efc_Pmo ----- MENKGVIQIIGPVVDVTFE-NE-LPRIYNALKIDR-GNG-EYLVAEVQ
 efc_Bsu ----- MKKGRVSQLVGPVVDVRFEDGHLPEIYNAIKISQPAASENEVGIDLTLLEV
 efc_Ehi ----- MSSGKIVQVIGPVVDVQFEDSLQSLPDINNALVYVKK--- NDEKSKSVVLETA
 efc_Sli ----- MTTTVEATATGRVWARVIGPVVDVQFEDPA-MPEIYNAHVLEVADPAKEGELKTTLLEV
 afc_Mba ----- MVDRSRLNLGKVISVRGSIVDVLF-EKHLPPVYTLRA--- GRESQIAIEV
 vvc_Ncr QNGA----- EVDOIHTGKJYSVSGPVVVAEDMIGVAMYELVKY--- GHDLQVGEV
 vvc_Sce IEENARKEIKRISLEDHAESEYGAISVSGPVVIAENMIGCAMYELVKY--- GHDNLVGEV
 vvc_Hvu ----- ELVRYV---
 vvc_Zma
 vvc_Bta ----- MDFSKLKPKIRDDEKESTFGYVHGVSGPVTACDMAGAACMYELV
 vvc_Dme ----- MSNLRKFKDEERESEYGRVYAVSGPVSAAEAMSGSAMYEELV
 vvc_Tco ----- MTSDKNPKYTEQRMGAVKAVSGPVIIAENMGGSAMYEELV
 avc_Mja ----- MKREVENMPVVGKIIKIAVGPPVVAEGMKGQAQMYEVVK
 evc_Ehi ----- MQIGKIKVSGPLVMAENMSEASIQDMCLV
 avc_Mba ----- MEVKGEIYRVSGPVTIAIG-LQAKMYDLVK
 avc_Mma ----- MEVKGEIYRVAGPVTIAIG-LDAKMYDLCKV
 avc_Hsa ----- MSQA-EAITDTGEIESVSGPVTATG-LDAQMNDVYV
 avc_Hvo ----- MSQATQDSVYREDGVIASVSGPVTARG-LDARMNDVYV
 evc_Tth ----- MIQGVQVQTAGPAVIAKGMLGARMYDICKV
 avc_Sac ----- MVSEGRVVRVNGPLVIADGMREAQMFEV
 + + + + +

vvr_Ncr E ARGNRAV VQVFEGTSGIDVKKTKEFTGESLKLGVSEDMLGRIFDGSGRAIDKGPKVLAE-----
 vvr_Sce E IRGDRAIVQLFEGTSGIDVKTTVEFTGESLRIPVSEDMGLRIFDGSGRPIDNGPKVFAE-----
 vvr_Bta EVSGSKAVVQVFEGTSGIDAKKTSCEFTGDIILRTPVSEDMGLGRVFNGSGKPIDRGPVVLAE-----
 vvr_Dme EVSGSKAVVQVFEGTSGIDAKNTLCFTGDIILRTPVSEDMGLGRVFNGSGKPIDKGPPILAE-----
 vvr_Hvu EVDGEKAVVQVFEGTSGIDDNKYTTVQFTGEVLKTPVSLDMGLRIFNGSGKPIDNGPPILPE-----
 vvr_Tco EVGDGKAVVQVFEGTSGIDVMRSKCEFTGKVMELGVSEDMLGRVFNGSGIPIDDGPPVLAE-----
 avr_Mba DSSSDIVVIQIFEGTTGLD-KECGVVFTGETLKLPAASIDLLGRILSGSGEPLDGGPRIVPD-----
 avr_Mma DSLLTLWRSFRSKVLYGLT-RTAVVIFTGETLKLPAASV DLLGRILFGSGEPLDGGPRIVPD-----
 avr_Hsa ESSDFVIAQVVFEGTEGVG-KDASVRLGETLKMPVTEDLLGRVLDGSGNPIDGGPDIVPD-----
 avr_Hvo ESSEGGLVAIQVVFEGTSGID-RNASVRLGETLKMPVTEDLLGRVLDGSGQPIDGGPEIVPD-----
 avr_Mja EAREGLAVVQVFEGTTGLSTKDTRVRFTGRATAKIGVSMEMLGRIFNGAGKPIDGGPEIVPE-----
 evr_Ehi EVQEDKAMVQIFEGTSGINLNKNSSVRLGHPLQLGVSEDMIGR VF DGLGRPKDNGPEILPE-----
 evr_Tth EVSEEYAVIYQVFETTGLDLATT SVS WEDVARLGVSKEMLGR R FNGGKPIDGLPPITPE-----
 avr_Sac DSQMGTTFVQVFEGTTGISPTGSKVRLGRGLEVKISEMLGRIFNPLGEPLDNGPPVIGG-----
 cfr_Osa NLESKNVGIVLMDGLMIQ-EGSFVKATGRIAQIPVSEAYLGRVINALAKPIDGRGEIVAS-----
 cfr_Zma NLESKNVGIVLMDGLMIQ-EGSFVKATGRIAQIPVSEAYLGRVINALAKPIDGRGEIVAS-----
 cfr_Tae NLESKNVGIVLMDGLMIQ-EGSFVKATGRIAQIPVSEAYLGRVINALAKPIDGKGEIIAS-----
 cfr_Nta NLESNNYGVVLMDGLLIQ-EGSSVKATGRIAQIPVSEAYLGRVINALAKPIDGRGEISAS-----
 cfr_Sol NLESNNVGVVLMDGLMIQ-EGSSVKATGRIAQIPVSEAYLGRVINALAKPIDGRGEITAS-----
 cfr_Psa NLESKNVGIVLMDGLMIQ-EGSSVKATGRIAQIPVSEAYLGRVINALAKPIDGRGEISTS-----
 cfr_Mpo NLESNDNVGAVLMDGLTIQ-EGSSVKATGKIAQIPVSDAYLGRVINALAQPIDGKGQIPAS-----
 cfr_Cpa NLEADNVGAVLMDGRNIL-EGSSVRATQKIAQIPVSDAYLGRVVDALARPIDGKGDIATT-----
 cfr_Ppu NLESNDNVGAVLMDGRDIL-EGSSVKATGKIAQIPVSDAYLGRVVDPLARPIDNKGEASN-----
 efr_Syn NLEEDNVGIVLMDGLMIQ-EGSTVTRATGKIASVPGVGEAVIGRVRVOALMRPIDGKGEIHAT-----
 efr_Bsu NLEESNVGIVILGPFSER-EGDVKRTGRIMEVPGVGEELIGRIVNPLQGPVDGLGPILTS-----
 efr_Ehi NLESNDNVGIVILGDFESIR-EGDKVKRTGKIMEVPVGDA LIGR VVNPLQGPIDGLGEIVTD-----
 efr_Pmo NLEENNNGAVIILGDPGVK-E GDEVRATGQIAAVGAGEALLGRVNVSLGEPIDGKGELKTE-----
 mfr_Osa NLENENVGIVVFGSDTAIK-E GDLVKRTGSIVDVPGAKAMLGRVVDALGPVIDGKGALSDH-----
 mfr_Zma NLENENVGIVVFGSDTAIK-E GDLVKRTGSIVDVPGAKAMLGRVVDALGPVIDGKGALSDH-----
 mfr_Bta NLEPDNVGVVVFGNDKLIK-E GDIVKRTGAI DVDPVGEELLGRVVDALGNAIDGKGPIGSK-----
 mfr_Hsa NLEPDNVGVVVFGNDSLIH-E GDIVKRTGAI DVDPVGEELLGRVVDALGNAIDGKGPIGSK-----
 mfr_Dme NLEPDNVGVVVFGNDKLIK-QGDIVKRTGAI DVDPVGEELLGRVVDALGNAIDGKGAINTK-----
 mfr_Spu NLEPDNVGVVVFGNDKLIK-E GDVVKRTGSIVDVPGAKAMLGRVVDALGPVIDGKGPLGTE-----
 mfr_Ncr NLEAGQVGVVLFGSDRLVK-E GETVKRTGTEIVDGPVGEELLGRVVIDALGPVIDGKGPINCK-----
 efr_Rbl NLEVDNVGIVIFGDDRDIK-EGDTVKRTKSIVDVPA GNGLLGRVVDALGPVIDGKGPIVAS-----
 mfr_Cre NLQADHYGVVVFGNDSLIH-QGDLVYRTGQIVNVPGP GTLGRVVDALGPVIDGKGPLTNV-----
 efr_Sli NLEEREIGCVVLFGEFSIE-E GQPVPSRTGEVLSVAVAEGYLGRVVDPLGPVIDGKGPIEITS-----
 efr_Eco NLERDSVGA VVMPYADLA-E GMKVKTGRIEVPGVGRLLGRVWN T LGA P IDGKGPLHD-----
 afr_Mba NVDKEIGTVLLGEYSHLH-AGDQVERTRRVMDVAG EELLGRVVIDPLGPVIDNKGPIVSS-----
 cfc_Osa QLLGNNRVR A VAMSATDGLMRGM EVIDTGAPLSVPVGATLGRIFFNVLGEPVDNLGPVDT S-----
 cfc_Zma QLLGNNRVR A VAMSATDGLMRGM EVIDTGAPLSVPVGATLGRIFFNVLGEPVDNLGPVDT S-----
 cfc_Tae QLLGNNRVR A VAMSATDGLMRGM EVIDTGAPLSVPVGATLGRIFFNVLGEPVDNLGPVDT S-----
 cfc_Nta QLLGNNRVR A VAMSATDGLTRGM EVIDTGAPLSVPVGATLGRIFFNVLGEPVDNLGPVDT S-----
 cfc_Sol QLLGNNRVR A VAMSATDGLKRGMEV IDTGAPLSVPVGATLGRIFFNVLGEPVDNLGPVDT R-----
 cfc_Psa QLLGNNRVR A VAMSATDGLKRGMEV IDTGAPLSVPVGATLGRIFFNVLGEPVDNLGPVDT R-----
 cfc_Mpo QLLGNNKVR A VAMSATDGM MRGM KV IDTGAPLTVPVGATLGRIFFNVLGEPVDNLGPVDT-----
 cfc_Pth QLLGNHKVR A VAMSATDGLTRGM RV IDTGAPLSVPVGATLGRIFFNVLGEPVDNLGPVDA G-----
 cfc_Cpa QLLGDNQVR A VAMSSTIDGILRGMEV IDTGAPLSGPATLGRIFFNVLGEPVDELGAVVCD-----
 cfc_Pou QLLGDNQVR A VAMSSTIDGLVRGMEV IDTGAPLSGPATLGRIFFNVLGEPVDNLGPVSS-----
 efc_Syn QLLGDNQVR A VAMSSTIDGLVRGMEV IDTGAPLSVPVGATLGRIFFNVLGEPVDEKGEVNIS-----
 mfc_Osa QHLGENMVRTIAMDGTEGLVRGQRVLNTGSPIT TVPGVGRATLGRIMNVIGEPIDEKGDI TT-----
 mfc_Zma QHLGENMVRTIAMDGTEGLVRGQRVLNTGSPIT TVPGVGRATLGRIMNVIGEPIDEKGDI KT-----
 mfc_Nta QHLGENMVRTIAMDGTEGLVRGQRVLNTGSPIT TVPGVGRATLGRIMNVIGEPIDEKGDI NT-----
 mfc_Hbr QHMGEGMVRTIAMDGTEGLVRGQRVLNTGSPIT TVPGVGRANPWTYHEVIGEPIDEKGDI PITT-----
 mfc_Bta QHLGESTVRTIAMDGTEGLVRGQRVLNTGSPIT TVPGVGRANPWTYHEVIGEPIDEKGDI PI KTK-----
 mfc_Rno QHLGESTVRTIAMDGTEGLVRGQRVLNTGSPIT TVPGVGRANPWTYHEVIGEPIDEKGDI PI KTK-----
 mfc_Sce QHLGENTVRTIAMDGTEGLVRGQRVLNTGSPIT TVPGVGRANPWTYHEVIGEPIDEKGDI PI KTK-----
 efc_Rbl QHLGENTVRTCIAMDATEGLVRGAPVSDSGTIPSVPGNATLGRILNVVGEPEID ERGPIAAT-----
 mfc_Ncr QHLGENVVR CIAMDGTEGLVRGAKASDTGAPITIPVGPA TLGRILNVVGEPEID ERGPIKTD-----
 efc_Eco QOLGGGIVRTIAMGSSDGLRRLDVKDLEHPVGPATLGRILNVVGEPEID ERGPIKTD-----
 efc_Pmo QHLGNSVVR A VAMDATDGLQRGM EVVDTGPAITVPGKA V LGRILNVVGEPEID ERGPIKTD-----
 efc_Bsu LHLGDDTVRTIAMASTDGLQRGM EVVDTGPAISVPGDVTLGRVFNVLGENIDLNEPVPA DA-----
 efc_Ehi LELGDGVIRTIAMESTDGLQRGM EVVDTGPAISVPGDVTLGRVFNVLGENIDLNEPVPA DA-----
 efc_Sli QHLGDGLVR TISMQPTDGLIRQAPVDTGAAISVPGDVTLGRVFNVLGENIDLNEPVPA DA-----
 afc_Mba TQLDAAHHVRGIALTPTBLARGMAV ETDGPKLAPVGREILSRMF DFGNTIDRRKP-PSDI-----
 vvc_Ncr RINGDQATIQV-YEETAGV MVGDPVLR TGKPLSVELGPGLLNNIYDGIQRP LEKIAEASNSIYIPRGIA T-----
 vvc_Sce RIDGDKATIQV-YEETAGLTVGDPVLR TGKPLSVELGPGLMETIYDGIQRP LKAIKEESQSIYIPRGID T-----
 vvc_Hvu RLEGDSATIQV-YEETAGLTVNDPVLRTKPKPLSCELGPGLGNIFDGIQRP LK TIAIKSRDVYI P RGVSV-----
 vvc_Zma ---ARATIQV-YEETAGLMVNDPVLRTKPKPLSCELGPGLGNIFDGIQRP LK TIAIKSGD YI P RGVSV-----
 vvc_Bta RLEGDMATIQV-YEETSGVSVGDPVLR TGKPLSVELCPGIMGAI FDGIQRP LSDISSQTQSIYIPRGVNV-----
 vvc_Dme RLEGDMATIQV-YEETSGVTVGDRVLRTGKPLSVELGPGLMSIFDGIQRP LR DIGVMTNSIYIPKGVNT-----
 vvc_Tco RLEGDTATIQV-YEETTGGLTVGDPVYCTGKPLSLELGPGLMSEIFDGIQRP LD TIYRMVENVFI P RGVQV-----
 avc_Mja QLHDDKAVIQV-YEETSGIKPGEPVGVTGAPLSVELGPGLMRLAMYDGIQRP LTAIB EKTGSIFIPRGV DV-----
 evc_Ehi EMRQDVASIQV-YEETSGIGPGEPVRSTGEALSVELGPGLMNIISQMF DGIQRP LD TFM E VTQSNFLGRGVQL-----
 avc_Mba QILGPKTI IQV-YEETAGIKPGEPCVSTGSSL SVELGPGLLSSSIYDGLQRP LHVLLER TG-GFIGRGV TA-----
 avc_Mma QIVGDKTI IQV-YEETGGVRPGEPCVTTGMSL AVELGPGLLSSSIYDGLQRP LHVLLER TG-GFIGRGV TA-----
 avc_Hsa EIEGDVTTIQV-YEETSGIGPGQPV DNTGEPLTVDLGPGLMDSIYDGVQRPLDVLEDEM G-AFLDRGVDA-----
 avc_Hvo EIEGDOLTTIQV-YEETSGVGPGEVSTGEPLTVDLGPGLMDAIYDGVQRPLDVLESKMDSAFLDRGVDA-----
 evc_Tth RLQGDTA FVQV-YEETSGLKVGEPVVSTGLPLA VELGPGLMNGIYDGIQRP LERIREKTGI-YITRGV VV-----
 avc_Sac RIEGDRAFIQV-YEETSDGKPKDVKYRSGAPLSVELGPGLIGKIYDGLQRP LDSIAK VNSPFVARGV SI-----

	+	+	+	+	+	+	+
vvr_Ncr	-	-	-	-	-	-	-
vvr_Sce	-	-	-	-	-	-	-
vvr_Bta	-	-	-	-	-	-	-
vvr_Dme	-	-	-	-	-	-	-
vvr_Hvu	-	-	-	-	-	-	-
vvr_Tco	-	-	-	-	-	-	-
avr_Mba	-	-	-	-	-	-	-
avr_Mma	-	-	-	-	-	-	-
avr_Hsa	-	-	-	-	-	-	-
avr_Hvo	-	-	-	-	-	-	-
avr_Mja	-	-	-	-	-	-	-
evr_Ehi	-	-	-	-	-	-	-
evr_Tth	-	-	-	-	-	-	-
avr_Sac	-	-	-	-	-	-	-
cfr_Osa	-	-	-	-	-	-	-
cfr_Zma	-	-	-	-	-	-	-
cfr_Tae	-	-	-	-	-	-	-
cfr_Nta	-	-	-	-	-	-	-
cfr_Sol	-	-	-	-	-	-	-
cfr_Psa	-	-	-	-	-	-	-
cfr_Mpo	-	-	-	-	-	-	-
cfr_Cpa	-	-	-	-	-	-	-
cfr_Ppu	-	-	-	-	-	-	-
efr_Syn	-	-	-	-	-	-	-
efr_Bsu	-	-	-	-	-	-	-
efr_Ehi	-	-	-	-	-	-	-
efr_Pmo	-	-	-	-	-	-	-
mfr_Osa	-	-	-	-	-	-	-
mfr_Zma	-	-	-	-	-	-	-
mfr_Bta	-	-	-	-	-	-	-
mfr_Hsa	-	-	-	-	-	-	-
mfr_Dme	-	-	-	-	-	-	-
mfr_Spu	-	-	-	-	-	-	-
mfr_Ncr	-	-	-	-	-	-	-
efr_Rbl	-	-	-	-	-	-	-
mfr_Cre	-	-	-	-	-	-	-
efr_Sli	-	-	-	-	-	-	-
efr_Eco	-	-	-	-	-	-	-
afr_Mba	-	-	-	-	-	-	-
cfc_Osa	-	-	-	-	-	-	-
cfc_Zma	-	-	-	-	-	-	-
cfc_Tae	-	-	-	-	-	-	-
cfc_Nta	-	-	-	-	-	-	-
cfc_Sol	-	-	-	-	-	-	-
cfc_Psa	-	-	-	-	-	-	-
cfc_Mpo	-	-	-	-	-	-	-
cfc_Pth	-	-	-	-	-	-	-
cfc_Cpa	-	-	-	-	-	-	-
cfc_Ppu	-	-	-	-	-	-	-
efc_Syn	-	-	-	-	-	-	-
mfc_Osa	-	-	-	-	-	-	-
mfc_Zma	-	-	-	-	-	-	-
mfc_Nta	-	-	-	-	-	-	-
mfc_Hbr	-	-	-	-	-	-	-
mfc_Bta	-	-	-	-	-	-	-
mfc_Rna	-	-	-	-	-	-	-
mfc_Sce	-	-	-	-	-	-	-
efc_Rbl	-	-	-	-	-	-	-
mfc_Ncr	-	-	-	-	-	-	-
efc_Eco	-	-	-	-	-	-	-
efc_Pmo	-	-	-	-	-	-	-
efc_Bsu	-	-	-	-	-	-	-
efc_Ehi	-	-	-	-	-	-	-
efc_Sli	-	-	-	-	-	-	-
afc_Mba	-	-	-	-	-	-	-
vvv_Ncr	PALDRKKKWEFTPT--MKVGDHIAAGGDYWGTVYENSFISVHKILLPPRARGTITTRIAEKGEYTVEEKILE						
vvv_Sce	PALDRTIKWQFTPQPK-FQVGDHISGGDIYGSVFENSLISSHHKILLPPRSRGTTITWIAPAGEYTLDEKILE						
vvv_Hvu	PALDKDQLWEFQPNK-LGVGDNITNGDLYATVFENTLM-KHHIALPPGAMGKISIYIAPAGQYSLQDTVLE						
vvv_Zma	PALDKDVLWEFPKT-LGVGDVITGGDLYATVFENTLM-QHHVALPPGSMGKISIYIAPAGQYNLQDTVLE						
vvv_Bta	SALS RDVKWDFTPCKNLRVGSHITGGDIYGINNENSLI-KHKIMLPPRNRGTVTYIAPPGNYDTSDVVLE						
vvv_Dme	TALSRSEMWEFNP-LNVRVGSHTGGDLYGVVHENTLV-KQRMIVAPRAKGTVRYIAPAGNNLEDIVLE						
vvv_Tco	KSLNDQKQWDFKPC--LKVGDLVSGGDIIGSVENSLMYNSIMIPPNVRGRVTISIVPSGNYTLQDDIE						
avc_Mja	PALPRDIKWEFKPV--VNEGDYVEEGDIIGTVDETPSI-VHKILVPIGVK-GKIVEIKEGKFTVEETVAV						
evc_Ehi	PALDHEKQWWFEAT--IEEGTEVSAGDIIGGYVDETPKII-QHKIMVPNGIK-GTVQKIESGSFTIDDPICV						
avc_Mba	DGLDHKKLWEFKPI--VKKGDSVKGDGIVGVVQETVNI-EHKIMLPPDI-SGTVDIIRSGSFTVVDTICT						
avc_Mma	DGLDHKKLWEFKPV--VKKGDSVKGDGIVGVVQETVNI-EHKIMLPPDI-SGTVDIIRSGSFTVVDTICT						
avc_Hsa	PGIDLDTDWEFEPT--VEAGDEVAAGDVVGTVDETVSI-EHKVLVPPRSODGEVVAVESGFTVDDTVVE						
avc_Hvo	PGIDLDEKWEFEPT--VSEGDEVAPGDVVGTVPETVTI-EHKVMVPPDFGGGEVVAVEEGEFSVTEAVVE						
evc_Tth	HALDREKKWAUTPM--VKPGDEVRGGMVLGTVPE-FGF-THKILVPPDVGRVKEVKPAGEYTVEEPVVV						
avc_Sac	PALDRQTKWHFVPK--VKSGDKVGPQDIIGVVQETDLI-EHRILIPPNVHGTLELAREGDYTVEDVVV						

+ + + + +
 vvr_Ncr ----- EYLDINGSPINPYAREYPQE MISTGISAIDTMNSIARGQKIPIFSAAGLPHNEIAA
 vvr_Sce ----- DYLDINGSPINPYARIYPEEMISTGVSAILDMNSIARGQKIPIFSAAGLPHNEIAA
 vvr_Bta ----- DFLDIMGQPINPHFRIYPEEMIQTGISAIDGMNSIARGQKIPIFSAAGLPHNEIAA
 vvr_Dme ----- DFLDIQGQPINPWSRIYPEEMIQTGISAIDVMNSIARGQKIPIFSAAGLPHNEIAA
 vvr_Hvu ----- AYLDISGSSINPSERTYPEEMIQTGISTIDVMNSIARGQKIPPLFSAGLPHNEIAA
 vvr_Tco ----- QYRDIQGQIPINPRARVYYPEEMIQTGISSIDVMTSISRGQKIPPLFSAGLPHNEIAA
 avr_Mba ----- QLLDINGAAAMNPYARLPPKDFIQTGISTIDGNTLVRGQKLPPIFSASGLPHNEIAL
 avr_Mma ----- QLLDINGAAAMNPYARLPPKDFIQTGISTIDGNTLVRGQKLPPIFSASGLPHNEIAL
 avr_Hsa ----- DRDIVGEAINPHAREYPEEFIQTGVSGIDGMNTLVRGQKLPPIFSGSGLPHSDLAL
 avr_Hvo ----- ERDIVGAAINPYSREYPEEFIQTGVSAIDGMNTLVRGQKLPPIFSASGLPHNDLAL
 avr_Mja ----- NELDINGYPLNPYRSRKVPSDFIQTGISTIDGMNTLVRGQNLPIFSGSGLPHNQLAA
 evr_Ehi ----- KYLDINGEVINPIARDPDEFIQTGISAIDHLNTLVRGQKLPVFSGSGLPHKELA
 evr_Tth ----- KRLPITGLPLNPVARRKPEQFIQTGISTIDGMNTLVRGQKLPPIFSGSGLPHNEIAA
 avr_Sac ----- ENRNINGDPINPATREYPEEFIQTGISAIDGLNSLLRGSKITDLSGSGLPANTLAA
 cfr_Osa ----- ESRLIESPAPGIISRRSVYEPLQTGLIAIDSMPIGRGRQRELIGDRQTGKTAVAT
 cfr_Zma ----- ESRLIESPAPGIISRRSVYEPLQTGLIAIDSMPIGRGRQRELIGDRQTGKTAVAT
 cfr_Tae ----- ESRLIESPAPSIIISRRSVYEPLQTGLIAIDSMPIGRGRQRELIGDRQTGKTAVAT
 cfr_Nta ----- EFRLIESAAGGIISRRSVYEPLQTGLIAIDSMPIGRGRQRELIGDRQTGKTAVAT
 cfr_Sol ----- ESRLIESPAPGIMSRSSVYEPLQTGLIAIDSAMIPVGRGRQRELIGDRQTGKTAVAT
 cfr_Psa ----- ESRLIESPAPGIISRRSVYEPLQTGLIAIDSMPIGRGRQRELIGDRQTGKTAVAT
 cfr_Mpo ----- EFRLIESPAPGIISRRSVYEPMQQTGLIAIDSMPIGRGRQRELIGDRQTGKTAVAI
 cfr_Cpa ----- DTRLIESSAPGIISRKFYVEPLQTGITAIDAMIPIPRGRQRELIGDRQTGKTAVAI
 cfr_Ppu ----- GTRLIESMAPGIVIGRQSVCEPMQQTGITAIDSMPIGRGRQRELIGDRQTGKTAVAI
 efr_Syn ----- ATRLIESPAPGIIVRKSCVEPLQTGITAIDAMIPIGRGRQRELIGDRQTGKTAVAI
 efr_Bsu ----- NTRPIESPAPGVMDRKSVHEPLQTGINKAIDALIPIGRGRQRELIGDRQTGKTSAI
 efr_Ehi ----- NVRPVEAMAPGVMDRKSVNEPMQQTGLKAIDALVPIGRGRQRELIGDRKTGKTSAI
 efr_Pmo ----- NMMLDRLKAYGIIISRKPVHEPLQTGKSIDGMVPIGRGRQRELIGDRQTGKTAVAL
 mfr_Osa ----- ERRRVEVKAPGIIERKSVHEPMQQTGLKAVDSLVPGRGRQRELIGDRQTGKTAVAI
 mfr_Zma ----- ERRRVEVKAPGIIERKSVHEPMQQTGLKAVDSLVPGRGRQRELIGDRQTGKTAVAI
 mfr_Bta ----- ARRRVGLKAPGIIPRISVREPMQQTGKAVDSLVPGRGRQRELIGDRQTGKTAVAI
 mfr_Hsa ----- TRRRVGLKAPGIIPRISVREPMQQTGKAVDSLVPGRGRQRELIGDRQTGKTSAI
 mfr_Dme ----- DRFRVGKAPGIIPRVSVERPMQQTGKAVDSLVPGRGRQRELIGDRQTGKTALAI
 mfr_Spu ----- HRARVGTAKPG-IPIRISVREPMQQTGKAVDSLVPGRGRQRELIGDRQTGKTAVAI
 mfr_Ncr ----- ENRAQLKAPGILPRQSVNQPVQTGKNSVDAMVPIGRGRQRELIGDRQTGKTAVAI
 efr_Rbl ----- ERRIADVKAPGIIPRKGVHEPMATGLKAVDAMIPIGRGRQRELIGDRQTGKTAVAI
 mfr_Cre ----- RSSLVEIKAPGIARQSVREPLYTGVKAVALVPIGRGRQRELIGDRQTGKTAVAL
 efr_Sli ----- GRRALELQAPTVMORKSVHEPMETGYKAVDAMTPIGRGRQRLIIGDRQTGKTALAV
 efr_Eco ----- GFSAVEAIAPGVIERQSVSDQPVQTGYKAVDSMIPIGRGRQRELIGDRQTGKTALAI
 afr_Mba ----- NRPIERPSPAIMDRAPVTVPLOTVKVIDALIPIGRGRQRELILGDRQTGKTAVAI
 cfc_Osa ----- ATFPIHRSAPAFIELDTCKLISIFETGKVVDDLLAPYRGGKIGLFGGAGVGKTVLIM
 cfc_Zma ----- ATFPIHRSAPAFIELDTCKLISIFETGKVVDDLLAPYRGGKIGLFGGAGVGKTVLIM
 cfc_Tae ----- ATFPIHRSAPAFIELDTCKLISIFETGKVVDDLLAPYRGGKIGLFGGAGVGKTVLIM
 cfc_Nta ----- TTSPIHRSAPAFIQLDTCKLISIFETGKVVDDLLAPYRGGKIGLFGGAGVGKTVLIM
 cfc_Sol ----- TTSPIHRSAPAFIQLDTCKLISIFETGKVVDDLLAPYRGGKIGLFGGAGVGKTVLIM
 cfc_Psa ----- TTSPIHRSAPAFIQLDTCKLISIFETGKVVDDLLAPYRGGKIGLFGGAGVGKTVLIM
 cfc_Mpo ----- TTFPIHRAAAPAFTOLDTCKLISIFETGKVVDDLLAPYRGGKIGLFGGAGVGKTVLIM
 cfc_Pth ----- ITSPIHRPAPAFTELDTCKLISIFETGKVVDDLLAPYRGGKIGLFGGAGVGKTVLIM
 cfc_Cpa ----- STLPIHRSAPAFQOLETKSSIFETGKVVDDLLAPYRGGKIGLFGGAGVGKTVLIM
 cfc_Ppu ----- STLPIHRSAPAFPTKLETKPSIFETGKVVDDLLAPYRGGKIGLFGGAGVGKTVLIM
 efc_Syn ----- ETLPiHRPAPSFTELTKPSVETGKIVIDLLTPYRGGKIGLFGGAGVGKTVIMM
 mfc_Osa ----- HFLPIHREAPAFVEQATEQQILVTGKVVVDLYVAPYQRGGKIGLFGGAGVGKTVLIM
 mfc_Zma ----- HFLPIHREAPAFVEQATEQQILVTGKVVVDLLAPYQRGGKIGLFGGAGVGKTVLIM
 mfc_Nta ----- HFLPIHREAPAFVEQATEQQILVTGKVVVDLLAPYQRGGKIGLFGGAGVGKTVLIM
 mfc_Hbr ----- HFLPIHREAPAFVQDATEQQILVTGKVVVDLLAPYQRGGKIGLFGGAGVGKTVLIM
 mfc_Bta ----- QFAAIHAEAPEFVEMSVQEIEILVTGKVVVDLLAPYAKGGKIGLFGGAGVGKTVLIM
 mfc_Rno ----- QFAPIHAEAPEFVIEMSVQEIEILVTGKVVVDLLAPYAKGGKIGLFGGAGVGKTVLIM
 mfc_Sce ----- LRKPIHADPPSFAEQSTSIAEILETGKVVVDLLAPYARGGKIGLFGGAGVGKTVFIQ
 efc_Rbl ----- ENRAIHQKAPDFQSQSTESQILVTGKIVIDLLAPYSKGGKIGLFGGAGVGKTVLIQ
 mfc_Ncr ----- NFRIPIHAEAPEFVEQSTTAEILVTGKIVVVDDLLAPYARGGKIGLFGGAGVGKTVFIQ
 efc_Eco ----- ERWAIHRAAPSYYELNSQSELLETGKIVIDLMCPFAKGKVKLFGGAGVGKTVNMHM
 efc_Pmo ----- EYAPIHREAPAFEDQGTEKEVFTETGKIVVVDDLLAPYVKGKIGLFGGAGVGKTVLIM
 efc_Bsu ----- KNDPIHQRAPSFDQLSTEVEILETGKIVVVDDLLAPYINGKIGLFGGAGVGKTVLIO
 efc_Ehi ----- ERSEIHKKAPSFDLSTSTEILETGKIVIDLLAPYINGKIGLFGGAGVGKTVLIO
 efc_Sli ----- ERWPIHRKAPNFDELESKTEMETGKIVIDLLTPYVKGKIGLFGGAGVGKTVLIO
 afc_Mba ----- QWRSIHQTPPPLMRRSTTSEIETGKIVIDLVPLERGGKAGLFGGAGVGKTVLIT
 vvc_Ncr ----- VEFDG-KKTEYPMQMTWPVRVPRPAEKHSANQPFV-YGQRVLDALFPSSVQGGTVIAIPGAFGCCKTVISQ
 vvc_Sce ----- VEFDG-KKSDFTLYHTWPVRVPRPVTekLsADYPLL-TGQRVLDALFPVCVQGGTCIPGAFGCCKTVISQ
 vvc_Hvu ----- LEFQG-IKKEFTMLHTWPVRTPRPRVASKLAADTPL-TGQRVLDALFPSSVLLGGTCAIIPGAFGCCKTVISQ
 vvc_Zma ----- LEFQG-IKKFTMLQTWPVRSPRPRVASKLAADTPL-TGQRVLDALFPSSVLLGGTCAIIPGAFGCCKTVISQ
 vvc_Bta ----- LEFEG-IKEKFSMVQWPVRQVRPVTekLpanhPLL-TGQRVLDALFPVCVQGGTTAIPGAFGCCKTVISQ
 vvc_Dme ----- TEFDG-EITKHTMLQWPVRHARPVTEkLpanhPLL-TGQRVLDSLFPVCVQGGTTAIPGAFGCCKTVISQ
 vvc_Tco ----- LEYNG-TVKSLLMHRWPVRTPRPRVASKESGNPLL-TGQRVLDALFPSSVQGGTCAIIPGAFGCCKTVISQ
 avc_Mja ----- VETENGERKEITMMQKWPVRKPRPYKEKLPPPEIPL-TGQRVEDTFFTLAKGGTAAIPGPFGSGKTVTQH
 evc_Ehi ----- IETEQGL-KELTMQMWPVRGRPIKQLNPDVPMI-TGQRVIDTFFPLAKGGTAAIPGPFGSGKTVTQH
 avc_Mba ----- LTDTG---ELQMMQRWPVRPRPVPKAKLTPTRPLV-TGMRILDGLFPVAKGGTAAIPGPFGSGKTVTQH
 avc_Mma ----- LTEGP---ELQMMQRWPVRKPRPVKRKLTPEKPLV-TGQRILDGLFPVAKGGTAAIPGPFGSGKTVTQH
 avc_Hsa ----- LDTEG---EIQMHQEWPRRQRPTVDKQTPTEPLV-SGQRILDGLFPPIAKGGTAAIPGPFGSGKTVTQH
 avc_Hvo ----- LDSGE---EITMHQEWPVRQARPAEKKTPREPLV-SGQRILDGLFPPIAKGGTAAIPGPFGSGKTVTQH
 evc_Tth ----- LEDGT---ELKMYHTWPVRARPVQRKLDPNTPFL-TGMRILDVLPVAMGGTAAIPGPFGSGKTVTQH
 avc_Sac ----- VDMNGDEI-PVKMYQNWPRIPRPPYKEKLEPVEPILL-TGIRVLDTVFPIAKGGTAAIPGPFGSGKTVTQH

+ + + + + + +
 vvr_Ncr QICRQAGLVRQRQGITNKGVHDGHEENFSIVFIGAMGVNLETARFFTRDSEE-----NGSLGRITLF
 vvr_Sce QICRQAGLVR---PTKDVHDGHEENFSIVFAAMGVNLETARFFKQDFEE-----NGSLERTLSF
 vvr_Bta QICRQAGLYK-KSKD---VVDYSEENFAIVFAAMGVNMETARFFKSDFEE-----NGSMDNVCLF
 vvr_Dme QICRQAGLVLPGKS---VLDDHTDNFAIVFAAMGVNMETARFFKQDFEE-----NGSMENVCLF
 vvr_Hvu QICRQAGLVKRLEQSKHAAEGGEEDNFAIVFAAMGVNMETAQFFKRDSEE-----NGSMERYVLF
 vvr_Tco QIVRQAGLYRK-----EGKQEDFCIVFAAMGVNQETARFFRTFEE-----NGSMEKT1VF
 avr_Mba QIARQAAVP-----GSESAFAVFAAMGITNEEAQYFMSDFEK-----TGALERAVVF
 avr_Mma QIARQASVP-----GSESAFAVFAAMGITNEEAQYFMSDFEK-----TGALERAVVF
 avr_Hsa QIARQASVPE----EEAETDDDEGESEFAVFGAMGITAAEANEFMDFER-----TGALERSVVF
 avr_Hvo QIARQATVPE----DE--ESGE[ESEFAVFGAMGITQEEANEFMDFER-----TGALERSVVF
 avr_Mja QIARQAKVR-----GEGEKFAVFAAMGITSEANFFMEFRK-----TGALERAVVF
 evr_Ehi QIARQATVL-----DSSDFAYVFAAIGITFEEAEFFMEDFRQ-----TGAIIDRSYMF
 evr_Tth QIARQATVRP---D-LSGEGEREPEFVFAAMGITQRELFSYFQEFR-----TGALSRSVLF
 avr_Sac QIAKQATVR-----GEESNFAVFAAIGVRYDEALFFRKFFEE-----TGAINRVAMF
 cfr_Osa DTILNQK-----QGDVICVYVAIGQRASSVAQVVTTFHE-----EGAMEYTTIV
 cfr_Zma DTILNQK-----QGDVICVYVAIGQRASSVAQVVTTFHE-----EGAMEYTTIV
 cfr_Tae DTILNQK-----QGDVICVYVAIGQRASSVAQVVTTFHE-----EGAMEYTTIV
 cfr_Nta DTILNQK-----QGNVICVYVAIGQKASSYAQVVTTLQE-----RGAMEYTTIV
 cfr_Sol DTILNQK-----QGNVICVYVAIGQKASSYAQVVTNFQE-----RGAMEYTTIV
 cfr_Psa DTILNQK-----QGNLVCVYVAIGQKASSYAQVVTTLQE-----RGAMEYTTIV
 cfr_Mpo DTILNQK-----QGNVVCVYVAIGQKASSYAQVVNTE-----RGALEYTTIV
 cfr_Cpa DTILNQK-----QGDVVVCVYVAIGQKASSYAQVVGVLQE-----KGALDYTTIV
 cfr_Ppu DTIINQK-----QGDVVVCVYVAIGQKASSYAQVVSLLQE-----KGALDYTTIV
 efr_Syn DTILNQK-----QGDVICVYVAIGQKASSYAQVNVNLRE-----RGALDYTTIV
 efr_Bsu DAIINQK-----DQDMICVYVAIGQKESTVRGVVETLRK-----HGALDYTTIV
 efr_Ehi DTILNQK-----QGDMICIYVAIGQKDTSTVRTQVETLKK-----YGAMDYTTIV
 efr_Pmo DAIINQK-----DTGVKCIYVAIGQKRSTVQAQIVKRLD-----AGALEYTTIV
 mfr_Osa DTILNQKQM-----NS-RGTNESETLYCVYVAIGQKRSTVQAQIVKRLD-----ANALEYSILV
 mfr_Zma DTILNQKQM-----NS-RGTNESETLYCVYVAIGQKRSTVQAQIVKRLD-----ANALEYSMLV
 mfr_Bta DTIINQKRF-----ND--GTDEKKKLYCIYVAIGQKRSTVQAQIVKRLTD-----ADAMKYTTIV
 mfr_Hsa DTIINQKRF-----ND--GSDEKKKLYCIYVAIGQKRSTVQAQIVKRLTD-----ADAMKYTTIV
 mfr_Dme DTIINQKRF-----ND--AQDESKKLYCIYVAIGQKRSTVQAQIVKRLTD-----SGAMGYSVIV
 mfr_Spu DTIINQKRF-----ND--EAEEKKNKLLYCIYVAIGQKRSTVQAQIVKRLHD-----SDALKYTII
 mfr_Ncr DAIINQKRW-----N-SGSDEOKKLYCVYVAVGQKRSTVQAQIVKRLHD-----NDAMKYSIV
 efr_Rbl DTILNQKVY-----NDAAGDDESKKLYCVYVAIGQKRSTVQAQIVKLEE-----TGAMAYTTIV
 mfr_Cre DCILHQNYL-----NGL-TNKKNRVYCIVYVAIGQKRSTVANLVKLFQ-----TGALKYTII
 efr_Sli DTIINQR-----DNWRTGDPNQVRCIYVAIGQKGSTIASVRGALEE-----NGALEYTTIV
 efr_Eco DAIINQR-----DSGIKCIYVAIGQKASTISNVVRKLEE-----HGALANTIV
 afr_Mba DTILNQRNF-----NVLCVYCAIGQRASAYARAVANLRE-----RGAMDYTTIV
 cfc_Osa ELINN-----IAKAHGGVSFVFGGVGERTREGNDLYMEMKE--SGVIN-EKNLEESKVALV
 cfc_Zma ELINN-----IAKAHGGVSFVFGGVGERTREGNDLYMEMKE--SGVIN-EKNIEESKVALV
 cfc_Tae ELINN-----IAKAHGGVSFVFGGVGERTREGNDLYMEMKE--SGVIN-EKNIEESKVALV
 cfc_Nta ELINN-----IAKAHGGVSFVFGGVGERTREGNDLYMEMKE--SGVIN-EENIAESKVALV
 cfc_Sol ELINN-----IAKAHGGVSFVFGGVGERTREGNDLYMEMKE--SGVIN-EQNIAESKVALV
 cfc_Psa ELINN-----IAKAHGGVSFVFGGVGERTREGNDLYMEMKE--SGVIN-EKNAESKVALV
 cfc_Mpo ELINN-----ILKAHGGVSFVFGGVGERTREGNDLYMEMKE--SGVIN-EQNISESKVALV
 cfc_Pth ELINN-----IAKAHGGVSFVFGGVGERTREGNDLYMEMKE--SGVIN-EQKISESKVALV
 cfc_Cpa ELINN-----IAKAHGGVSFVFGGVGERTREGNDLYMEMKE--SGVIN-ENNLPASKVALV
 cfc_Ppu ELINN-----IAKAHGGVSFVFGGVGERTREGNDLYMEMKE--SGVIN-EDNLKESKVALV
 efc_Syn ELINN-----IATQHGGVSFVFGVGETREGNDLYMEMIE--SGVID-KDOPSKSKIALV
 mfc_Osa ELINN-----VAKAHGGFSVFAVGERTREGNDLYREMIE--SGVIKLGDKQSESKCALV
 mfc_Zma ELINN-----VAKAHGGFSVFAVGERTREGNDLYREMIE--SGVIKLDKQSESKCALV
 mfc_Nta ELINN-----VAKAHGGFSVFAVGERTREGNDLYREMIE--SGVIKLGKQSESKCALV
 mfc_Hbr ELINN-----VAKAHGGFSVFAVGERTREGNDLYREMIE--SGVIKLGQADSKCALV
 mfc_Bta ELINN-----VAKAHGGFSVFAVGERTREGNDLYREMIE--SGVINLKD-ATSKVALV
 mfc_Rno ELINN-----VAKAHGGFSVFAVGERTREGNDLYREMIE--SGVINLKD-ATSKVALV
 mfc_Sce ELINN-----IAKAHGGFSVFTGVGERTREGNDLYREMIE--TGVINL-E-GESKVALV
 efc_Rbl ELINN-----IAKVHSGYSVFAVGERTREGNDLYHEFIE--SGVINIDD-LEKS^KVALV
 mfc_Ncr ELINN-----IAKAHGGFSVFTGVGERTREGNDLYHEMIE--SGVINLKD-ATSKVALV
 efc_Eco ELIRN-----IAIEHSGYSVFAVGERTREGNDLYHEMIE--T^KVID-^KVSLV
 efc_Pmo ELINN-----IAQGHGGGSVFAVGERTREGNDLYDEMIE--SGVLD-KTSLV
 efc_Bsu ELINN-----IAQEHHGGISVFAVGERTREGNDLYEMSD--SGVIN-KTAMV
 efc_Ehi ELIHN-----IAQEHHGGISVFTGVGERTREGNDLYEMKD--SGVIE-KTAMV
 EMIYR-----VANNHDGVSVFAVGERTREGNDLYEMKD--SGVID-KTALV
 afc_Mba EMIH^N-----VVKQHQGVSI^EFCG^ITGERCREGEELYRDMKD-AVL--PNTV^MV
 vvc_Ncr SVSKF-----SNSDVIVYYGCGERGNEMAEVLKDPEL^SIE-DGRKEPIMKRTT^L
 vvc_Sce SLSKY-----SNSDAIIYYGCGERGNEMAEVLM^EFPELYTEM-SGTKEPIMKRTT^L
 vvc_Hvu ALSKY-----SNSDTV^IYYGCGERGNEMAEVLMDFPQLTMTLPDGREESVMKRTT^L
 vvc_Zma ALSKY-----SNSEAV^IYYGCGERGNEMAEVLMDFPQLTMTLPDGREESVMKRTT^L
 vvc_Bta SLSKY-----SNSDVIIYYGCGERGNEMSEVLRDFPELT^ME-DGKVESIMKRT^T
 vvc_Dme ALSKY-----SNSDVIIYYGCGERGNEMSEVLRDFPELTCEI-DGVTESIMKRT^T
 vvc_Tco ALSKF-----SNSDAVIYYGCGERGNEMAEVLMDFPTLT-T^VIDGREESIMKRT^T
 avc_Mja QLAKW-----SDADVVVYIYGCGERGNEMTEVIEFPHLE--DIRTGKLMDR^T
 evc_Ehi QIAKW-----SDVDLVVYIYGCGERGNEMT^DVVN^EFPELI--DPNTGESLMERTV^L
 avc_Mba SLAKW-----SDTEIVVYIYGCGERGNEMADVLSEFPELE--DPQTGRPLMERTV^L
 avc_Mma QLSKW-----SDTEIVVYIYGCGERGNEMADVLWEFPELE--EPQTGRPLMERTV^L
 avc_Hsa SLAKF-----ADADI^VVYIYGCGERGNEMTEVIEDFPEL^P--DPQTGNPLMARTT^L
 avc_Hvo QLAKW-----ADADI^VVYIYGCGERGNEMTEVIEDFPELD--DPKTGNPLMARTC^L
 evc_Tth SLAKW-----SNADVVVYIYGCGERGNEMT^DVLYEFPELT--DPKTGGPLMHRTV^L
 avc_Sac SLAKW-----SAAKVVIYYGCGERGNEMT^DDELRSFPKLK--DPWTGKPLLLRT^T

vvr_Ncr LNLANDOPTIERIITPRLALTAAEYYAYQLEKHVLVILTDLSSYCDALREVSAAREEVPGRRGFPGYMYTD
 vvr_Sce LNLANDOPTIERIITPRLALTAAEFLAYQCEKHVLVILTDMSSYADALREVSAAREEVPGRRGFPGYMYTD
 vvr_Bta LNLANDOPTIERIITPRLALTAAEFLAYQCEKHVLVILTDMSSYAEALREVSAAREEVPGRRGFPGYMYTD
 vvr_Dme LNLANDOPTIERIITPRLALTAAEFLAYQCEKHVLVILTDMSSYAEALREVSAAREEVPGRRGFPGYMYTD
 vvr_Hvu LNLANDOPTIERIITPRLALTAAEFLAYCQKHLVILTDMSSYADALREVSAAREEVPGRRGFPGYMYTD
 vvr_Tco MNLANDOPTIERIITPRLALTAAEFLAYDCQKHLVILTDMSSYADALREVSAAREEVPGRRGFPGYMYTD
 avr_Mba LNLADDPAVERIVTPRMLATAEFLAYEHGHHVLVILTDITNYAEALRQMGAARNEIPGRRGYPGYMYTD
 avr_Mma LNLADDPAVERIVTPRMLATAEFLAYEHGHHVLVILTDITNYAEALRQMGAARNEIPGRRGYPGYMYTD
 avr_Hsa MNLADDPAVERIVTPRMLATAEFLAYEKFHDYHVLVILTDITNYCEALRQIGAAREEVPGRRGYPGYMYTD
 avr_Hvo TNLADDPAVERIVTPRMLATAEFLAYAFDKDYHVLVILTDITNYCEALREIGAAREEVPGRRGYPGYMYTD
 avr_Mja INLADDPAIERILTPRMLATAEFLAYEKFDMHVLVILTDITNYCEALREISAARNEVPGRRGYPGYMYTD
 evr_Ehi MNLANDOPTIERIATPRMLATAEFLAYEKGHHVLVILTDITNYAEALREISAARREVPGRRGYPGYLYTN
 evr_Tth LNKADDPTIEPILTTPRMLATAEFLAYEHDYHVLVILTDITNYCEALREIGAAREEIPGRRGYPGYMYTD
 avr_Sac VTLANDPPSLKILTPKTALTLAEYLAFEDKDYHVLVILTDITNYCEALRSASREEVPGRGGYPGYMYTD
 cfr_Osa AEMADSPATLQYLAPYTGAAALAEYFMYR-ERHTLIIYDOLSKQAQAYRQMSLLRRPPGREAYPGDVYFL
 cfr_Zma AEMADSPATLQYLAPYTGAAALAEYFMYR-ERHTLIIYDOLSKQAQAYRQMSLLRRPPGREAYPGDVYFL
 cfr_Iae AEMADSPATLQYLAPYTGAAALAEYFMYR-ERHTLIIYDOLSKQAQAYRQMSLLRRPPGREAYPGDVYFL
 cfr_Nta AETADSPATLQYLAPYTGAAALAEYFMYR-ERHTLIIYDOLSKQAQAYRQMSLLRRPPGREAYPGDVYFL
 cfr_Sol AETADSPATLQYLAPYTGAAALAEYFMYR-ERHTLIIYDOLSKQAQAYRQMSLLRRPPGREAYPGDVYFL
 cfr_Psa AETADSPATLQYLAPYTGAAALAEYFMYR-ERHTLIIYDOLSKQAQAYRQMSLLRRPPGREAYPGDVYFL
 cfr_Mpo AETADSPATLQYLAPYTGAAALAEYFMYR-KQHTLIIYDOLSKQAQAYRQMSLLRRPPGREAYPGDVYFL
 cfr_Cpa AANADDPATLQYLAPYTGASIAEYFMYR-GQHTLVIYDOLSKQAQAYRQMSLLRRPPGREAYPGDVYFL
 cfr_Ppu TANADSPATLQYLAPYTGASIAEYFMYR-GKHTLVIYDOLSKQAQAYRQMSLLRRPPGREAYPGDVYFL
 efr_Syn AANASDPAALQYLAPYTGASIAEYFMYQ-GKHTLVIYDOLSKQAQAYRQMSLLRRPPGREAYPGDVYFL
 efr_Bsu TASASQAPLLYLAPYAGVTMAEEFMYN-GKHVLVYDOLSKQAQAYRELSSLRRPPGREAFPGDVYFL
 efr_Ehi NAGASQAPLLYLAPYAGTAMGEFMYN-GKHVLIIYDOLSKQAQAYRELSSLRRPPGREAYPGDVYFL
 efr_Pmo AATASESAPLQYHMAPYTGSMGEYFMYR-MOK-GEHVLIVYDOLSKHAVAIREMSLLRRPPGREAFPGDVYFL
 mfr_Osa AATASDPAPLQFLAPYSGCAMGEYFRDN-GMHALIYDOLSKQAQAYRQMSLLRRPPGREAFPGDVYFL
 mfr_Zma AATASDPAPLQFLAPYSGCAMGEYFRDN-GMHALIYDOLSKQAQAYRQMSLLRRPPGREAFPGDVYFL
 mfr_Bta SATASDAAPLQYLAPYSGCSMGEYFRDN-GKHALIYDOLSKQAQAYRQMSLLRRPPGREAYPGDVYFL
 mfr_Hsa SATASDAAPLQYLAPYSGCSMGEYFRDN-GKHALIYDOLSKQAQAYRQMSLLRRPPGREAYPGDVYFL
 mfr_Dme SATASDAAPLQYLAPYSGCSMGEYFRDN-GKHALIYDOLSKQAQAYRQMSLLRRPPGREAYPGDVYFL
 mfr_Spu SATASDAAPLQYLAPYSGCSMGEYFRDN-GKHALIYDOLSKQAQAYRQMSLLRRPPGREAFPGDVYFL
 mfr_Ncr AATASEAPLQYLAPFTGACVGEYFRDN-GKHSLSVIFDDLTQAVAYRQMSLLRRPPGREAYPGDVYFL
 efr_Rbl AATASDPAPMQFLAPFSGTAGEFFRDN-GRHALMIYDOLSKQAQAYRQMSLLRRPPGREAYPGDVYFL
 mfr_Cre SATASDAAPLQFLAPYSGCSMGEYFRDT-GKHAVIIYDOLSKQSVAYRQMSLLRRPPGREAFPGDVYFL
 efr_Sli AAPASDPAGFKYLAPYTGSAIGQQWMYE-GKHVLIIYDOLSKQADAYRAVSLLRRPPGREAYPGDVYFL
 efr_Eco VATASESALQYLARMPVALMGEYFRDR-GEDALIYDOLSKQAVAYRQISLRRPPGREAFPGDVYFL
 afr_Mba VTEGNDPSGLIYITPYAATSIAEYFMES-GRDVLLIVYDOLTNHARAYRELSSLRRPPGREAYPGDIFYI
 cfc_Osa YGQMNEPPGARMRVGLTALTMAEYFRDVNKQDVLFFIDNIFRFVQAGSEVSALLGRMPSAVGYQPTLSTE
 cfc_Zma YGQMNEPPGARMRVGLTALTMAEYFRDVNKQDVLFFIDNIFRFVQAGSEVSALLGRMPSAVGYQPTLSTE
 cfc_Tae YGQMNEPPGARMRVGLTALTMAEYFRDVNKQDVLFFIDNIFRFVQAGSEVSALLGRMPSAVGYQPTLSTE
 cfc_Nta YGQMNEPPGARMRVGLTALTMAEYFRDVNEQDVLFFIDNIFRFVQAGSEVSALLGRMPSAVGYQPTLSTE
 cfc_Sol YGQMNEPPGARMRVGLTALTMAEYFRDVNEQDVLFFIDNIFRFVQAGSEVSALLGRMPSAVGYQPTLSTE
 cfc_Psa YGQMNEPPGARMRVGLTALTMAEYFRDVNEQDVLFFIDNIFRFVQAGSEVSALLGRMPSAVGYQPTLSTE
 cfc_Mpo YGQMNEPPGARMRVGLTALTMAEYFRDVNEQDVLFFIDNIFRFVQAGSEVSALLGRMPSAVGYQPTLSTE
 cfc_Pth YGQMNEPPGARMRVGLTALTMAEYFRDVNEQDVLFFIDNIFRFVQAGSEVSALLGRMPSAVGYQPTLSTE
 cfc_Cpa YGQMNEPPGARMRVVALTALTMAEYFRDVNEQDVLFFIDNIFRFVQAGSEVSALLGRMPSAVGYQPTLSTE
 cfc_Ppu YGQMNEPPGARMRVGLTALTMAEYFRDINKQDVLFFIDNIFRFVQAGSEVSALLGRMPSAVGYQPTLSTE
 efc_Syn YGQMNEPPGARMRVGLSGLTMAEYFRDVNKQDVLFFVDNIFRFVQAGSEVSALLGRMPSAVGYQPTLSTE
 mfc_Osa YGQMNEPPGARARVGLTGLTVAVHFRDAEGQDVLFFIDNIFRFQANSEVSALLGRIPSAVGYQPTLATE
 mfc_Zma YGQMNEPPGARARVGLTGLTVAEHFRDAEGQDVLFFIDNIFRFQANSEVSALLGRIPSAVGYQPTLATE
 mfc_Nta YGQMNEPPGARARVGLTGLTVAEHFRDAEGQDVLFFIDNIFRFQANSEVSALLGRIPSAVGYQPTLATE
 mfc_Hbr YGQMNEPPGARARVGLTGLTVAEHFRDAEGQDVLFFIDNIFRFQANSEVSALLGRIPSAVGYQPTLATE
 mfc_Bta YGQMNEPPGARARVALTGLTVAEYFRDQEKGQDVLFFIDNIFRFQAGSEVSALLGRIPSAVGYQPTLATE
 mfc_Rno YGQMNEPPGARARVALTGLTVAEYFRDQEKGQDVLFFIDNIFRFQAGSEVSALLGRIPSAVGYQPTLATE
 mfc_Sce FGQMNNEPPGARARVALTGLTIAEYFRDEEGQDVLFFIDNIFRFQAGSEVSALLGRIPSAVGYQPTLATE
 efc_Rbl FGQMNNEPPGARARVALTGLTIAEYFRDEEGQDVLFFIDNIFRFQAGSEVSALLGRIPSAVGYQPTLATE
 mfc_Ncr FGQMNNEPPGARARVALTGLTIAEYFRDEEGQDVLFFIDNIFRFQAGSEVSALLGRIPSAVGYQPTLATE
 efc_Eco YGQMNEPPGNRLRVALTGLTMAEKFRD-EGRDVLLFVDNIYRYTLAGTEVSALLGRMPSAVGYQPTLAEE
 efc_Pmo YGQMNEPPGARLRVGLTGLTMAENFRD-----KLAED
 efc_Bsu FGQMNNEPPGARMVALTGLTMAEHRFDVQGQDVLFFIDNIFRFQAGSEVSALLGRMPSAVGYQPTLATE
 efc_Ehi FGQMNNEPPGARMVALTGLTIAEYFRDVEKGQDVLFFIDNIFRFQAGSEVSALLGRMPSAVGYQPTLATE
 efc_Sli FGQMDDEPPGTRLRLVALAGLTMAEYFRDVKQDVLFFIDNIFRFQAGSEVSALLGRMPSAVGYQPTLATE
 afc_Mba FGQMNNEPPGARFRVGHALTMAEYFRDERRDVLLIDNIFRFIAGSEVSGLMGQMPSLRIGYQPTLATE
 vvc_Ncr ANTSNMPVAAREASIYTGITVAEYFRDQ-GMVAMMADSSSRWAEALREISGRLEMPADQGFPAYLGAK
 vvc_Sce ANTSNMPVAAREASIYTGITLAEYFRDQ-GKNSVSIADSSSRWAEALREISGRLEMPADQGFPAYLGAK
 vvc_Hvu ANTSNMPVAAREASIYTGITIAEYFRDM-GYNVNSMMADSTSRRWAEALREISGRLEMPADSGYPAYLASR
 vvc_Zma ANTSNMPVAAREASIYTGITIAEYFRDM-GYNVNSMMADSTSRRWAEALREISGRLEMPADSGYPAYLASR
 vvc_Bta ANTSNMPVAAREASIYTGITLSEYFRDM-GYHVNSMMADSTSRRWAEALREISGRLEMPADSGYPAYLGAR
 vvc_Dme ANTSNMPVAAREASIYTGITLSEYFRDM-GYNVAMMADSTSRRWAEALREISGRLEMPADSGYPAYLGAR
 vvc_Tco ANTSNMPVAAREASIYTGITLAEYFRDM-GKHIAAMMADSTSRRWAEALREISGRLEMPADGGGYPAYLASR
 avc_Mja ANTSNMPVAAREASVYTGITIAEYFRDM-GYGVLLTADSTSRRWAEAMREISGRLEMPGEEGYPAYLASR
 evc_Ehi ANTSNMPVAAREASIYTGITIAEYFRDM-GYDVAIMADSTSRRWAEALREMSGRLLEMPGDEGYPAYLGSR
 avc_Mba ANTSNMPVAAREASVYTGITIAEYFRDM-GLDVSLMADSTSRRWAEAMREISSSRLLEMPGEEGYPAYLASR
 avc_Mma ANTSNMPVAAREASVYTGMTIAEYFRDM-GYDVSLSMADSTSRRWPEAMREISSSRLLEMPGEEGYPAYLASR
 avc_Hsa ANTSNMPVAGRESCIYTGITIAEYFRDM-GYDVALMADSTSRRWAEAMREISSSRLLEMPGEEGYPAYLAAR
 avc_Hvo ANTSNMPVAARESCIYTGITIAEYFRDM-GYDVALMADSTSRRWAEAMREISSSRLLEMPGEEGYPAYLAAR
 evc_Tth ANTSNMPVAAREASIYVGTVTIAEYFRDQ-GFSVALMADSTSRRWAEALREISSRLLEMPAEEGYPYLAAR
 avc_Sac ANTSNMPVAARESSIYVGTVTMAEYFRDQ-GYDVLLVADSTSRRWAEALRDLGGRMEEMPAEEGFPYLAAR

vvr_Ncr LSTIYERAGRVAAGR----- + + + + + + + + + + + +
 vvr_Sce LSTIYERAGRVEGR----- - - - - NGSITQIPILTMPNDDITHPIPDLTGYITEGQIFVDRGLHNRG
 vvr_Bta LATIYERAGRVEGR----- - - - - NGSITQIPILTMPNDDITHPIPDLTGYITEGQIFVDRQLHNKG
 vvr_Dme LATIYERAGRVEGR----- - - - - NGSITQIPILTMPNDDITHPIPDLTGYITEGQIYVDRQLHNRQ
 vvr_Hvu LATIYERAGRRIEGR----- - - - - KGSITQIPILTMPNDDITHPTPDLTGYITEGQIYIDROLHNRQ
 vvr_Tco LACIYERAGRVLGR----- - - - - AGSITQIPILTMPNDDITHPIPDLTGYITEGQIYVDRQLHNRQ
 avr_Mba LATIYERAGIVKGA----- - - - - KGSVTQIPILSPMGDDITHPIPDLSGYITEGQIVVSRELHRKG
 avr_Mma LATIYERAGIVKGA----- - - - - KGSVTQIPILSPMGDDITHPIPDLSGYITEGQIVVARELHRKG
 avr_Hsa LAQLYERAGRIGEK----- - - - - EGSVTQIPILTMPGDDDTWPIPDLTGYITEGQIMMNRDLSNQG
 avr_Hvo LAQLYERAGRIGQR----- - - - - DGSVTQIPILTMPGDDDTWPIPDLTGYITEGQIYIDROLNSQG
 avr_Mja LATIYERAGRVKGR----- - - - - TGTITQIPILTMPDDITHPIPDLTGYITEGQIVLSRELHRKG
 evr_Ehi LATLFERAGRIRGL----- - - - - KGSVTQIPILTMPEDDKTHPIPDLTGYITEGQIILTRELYKG
 evr_Tth LATIYERAGVVEVGK----- - - - - KGSVTQIPILSPDDDTWPIPDLTGYITEGQIQLSRELHRKG
 avr_Sac LATIYERAGKVIGK----- - - - - KGSITQMPILTMPNDDDMTHPIPDLTGYITEGQIVLDRLFNKG
 cfr_Osa HSRLLERAALKNSLLGE----- - - - - GSMTALPIVETQSGDVSAIPTNVISITDGQIFLSADLFNAG
 cfr_Zma HSRLLERAALKNSLLGE----- - - - - GSMTALPIVETQSGDVSAIPTNVISITDGQIFLSADLFNAG
 cfr_Tae HSRLLERAALKNSLLGE----- - - - - GSMTALPIVETQSGDVSAIPTNVISITDGQIFLSADLFNAG
 cfr_Nta HSRLLERAALKSSLLGE----- - - - - GSMTALPIVETQSGDVSAIPTNVISITDGQIFLSADLFNAG
 cfr_Sol HSRLLERAALKSSLLGE----- - - - - GSMTALPIVETQSGDVSAIPTNVISITDGQIFLSADLFNAG
 cfr_Psa HSRLLERAVAKLSSQLGE----- - - - - GSMTALPIVETQSGDVSAIPTNVISITDGQIFLSADLFNAG
 cfr_Mpo HSRLLERAALKSSNLGE----- - - - - GSMTALPIVETQAGDVSAIPTNVISITDGQIFLSADLFNAG
 cfr_Cpa HSRLLERAALKSPQLGE----- - - - - GSMTALPIVETQAGDVCAYIPTNVISITDGQIFLSADLFNAG
 cfr_Ppu HSRLLERAALKNSDLGG----- - - - - GSMTALPIVETQAGDVSAIPTNVISITDGQIFLSADLFNAG
 efr_Syn HSRLLERAALKNDALGG----- - - - - GSMTALPIVETQAGDVSAIPTNVISITDGQIFLSADLFNAG
 efr_Bsu HSRLLERAALKSDAKGA----- - - - - GSITALPFVETQAGDISAYIPTNVISITDGQIFLQSOLFFSG
 efr_Ehi HSRLLERAALKSDDLGG----- - - - - GSITALPFVETQAGDISAYIPTNVISITDGQIFLESDLFYAG
 efr_Pmo HSRLLERAALKSDEIGA----- - - - - GSITALPIIETQAGDVSAIPTNVISITDGQIFLDSQLFNAG
 mfr_Osa HSRLLERAAKRSQDTGA----- - - - - GSITALPVIETQAGDVSAIPTNVISITDGQICLETLYFVRG
 mfr_Zma HSRLLERAAKRSQDTGA----- - - - - GSITALPVIETQAGDVSAIPTNVISITDGQICLETLYFVRG
 mfr_Bta HSRLLERAAKMNDAFGG----- - - - - GSITALPVIETQAGDVSAIPTNVISITDGQIFLETELIFYKG
 mfr_Hsa HSRLLERAAKMNDAFGG----- - - - - GSITALPVIETQAGDVSAIPTNVISITDGQIFLETELIFYKG
 mfr_Dme HSRLLERAAKMSPAMGG----- - - - - GSITALPVIETQAGDVSAIPTNVISITDGQIFLETELIFYKG
 mfr_Spu HSRLLERAAKMNPDFGG----- - - - - GSITALPVIETQAGDVSAIPTNVISITDGQIFLETELIFYKG
 mfr_Ncr HSRLLERAAKMNKTHGG----- - - - - GSITALPVIETQGGDVSAIPTNVISITDGQIFLESELIFYKG
 efr_Rbl HSRLLERSAKMNGDFA----- - - - - GSITALPVIETQAGDISAYIPTNVISITDGQIFLETELIFYKG
 mfr_Cre HSRLLERAALKSAAMGG----- - - - - GSITALFPVIETQAGDVSAIATNVISITDGQIFLETELIFYKG
 efr_Sli HSRLLERCAKSLDAEGA----- - - - - GSMTGLPIVETKANDVSASFIPTNVISITDGQCFLESOLFNAG
 efr_Eco HSRLLERAARVNAEYVEAFTKGEVKGKT----- - - - - GSITALPIIETQAGDVSAFVPNTNVISITDGQIFLETNLFNAG
 afr_Mba HSRLLERESTHLLKEFGG----- - - - - GSITALPVIETEAQNISAYIPTNLSITDGQIYLSPSLFELG
 cfc_Osa MGSLQERIT----- S - - - - TKKGSITSIQAIVYPADDLTDPAPATTFAHLDATTVLSRGLASKG
 cfc_Zma MGSLQERIT----- S - - - - TKKGSITSIQAIVYPADDLTDPAPATTFAHLDATTVLSRGLASKG
 cfc_Tae MGSLQERIA----- S - - - - TKKGSITSIQAIVYPADDLTDPAPATTFAHLDATTVLSRGLASKG
 cfc_Nta MGSLQERIT----- S - - - - TKEGSITSIQAIVYPADDLTDPAPATTFAHLDATTVLSRGLAAKG
 cfc_Sol MGSLQERIT----- S - - - - TKEGSITSIQAIVYPADDLTDPAPATTFAHLDATTVLSRGLAAKG
 cfc_Psa MGTLQERIT----- S - - - - TKEGSITSIQAIVYPADDLTDPAPATTFAHLDATTVLSRGLAAKG
 cfc_Mpo MGTLQERIT----- S - - - - TKEGSITSIQAIVYPADDLTDPAPATTFAHLDATTVLSRGLAAKG
 cfc_Pth MGSLQERIT----- S - - - - TKKGSITSIQAIVYPADDLTDPAPATTFAHLDATTVLSRGLAAKG
 cfc_Cpa MGSLQERIT----- S - - - - TTKGSITSIQAIVYPADDLTDPAPATTFAHLDATTVLSRNLAAKG
 cfc_Ppu MGALQERIT----- S - - - - TTEGSITSIQAIVYPADDLTDPAPATTFAHLDATTVLSRNLAAKG
 efc_Syn VGALQERIT----- S - - - - TMEGSITSIQAIVYPADDLTDPAPATTFAHLDATTVLSRNLAAKG
 mfc_Osa LGGLQERIT----- T - - - - TKKGSITSVQAIYVPADDLTDPAPATTFAHLDATTVLSRQISELG
 mfc_Zma LGGLQERIT----- T - - - - TKKGSITSVQAIYVPADDLTDPAPATTFAHLDATTVLSRQISELG
 mfc_Nta LGGLQERIT----- T - - - - TKKGSITSVQAIYVPADDLTDPAPATTFAHLDATTVLSRQISELG
 mfc_Hbr LGGLQERIT----- T - - - - TKKGSITSVQAIYVPADDLTDPAPATTFAHLDATTVLSRQISELG
 mfc_Bta MGTMQERIT----- T - - - - TKKGSITSVQAIYVPADDLTDPAPATTFAHLDATTVLSRAIAELG
 mfc_Rno MGTMQERIT----- T - - - - TKKGSITSVQAIYVPADDLTDPAPATTFAHLDATTVLSRAIAELG
 mfc_Sce MGLLQERIT----- T - - - - TKKGSVTSVQAVYVPADDLTDPAPATTFAHLDATTVLSRGISELG
 efc_Rbl MGQLQERIT----- S - - - - TKGGSITSVQAIYVPADDLTDPAPATSFAHLDATTVLSRAISELG
 mfc_Ncr MGQHQERIT----- T - - - - TKGGSITSVQAVYVPADDLTDPAPATTFAHLDATTVLSRGISELG
 efc_Eco MGVLQERIT----- S - - - - TKTGSITSVQAVYVPADDLTDPSPATTFAHLDATTVLSRGISELG
 efc_Pmo MGALQERIT----- S - - - - TKIGSITSVQAVYVPADDLTDPAPATTFHLDATTVLSRRIASLG
 efc_Bsu MGQLQERIT----- S - - - - TNVGSVTSIQAIVYPADDYTDPAPATTFAHLDATTNLERKLTEMG
 efc_Ehi MGQLQERIT----- S - - - - TKGGSITSIQAIVYPADDYTDPAPATTFAHLDATTVLSRQISELG
 efc_Sli MGLLQERIT----- S - - - - TRGHGSITSMQAIVYPADDLTUPAPATTFAHLDATTVLSRQISELG
 afc_Mba LSELEERIS----- T - - - - TDAAGAIMSIQAIVYPADDFTDPSAVHTFSHLASIVLSRKRASEG
 vvc_Ncr LASFYERAGRVKQALGSP----- - - - - PREGSVSVIVGAVSPPGDFSDPVTSAFLGIVQVFWGLDKKLAQRK
 vvc_Sce LASFYERAGRKAVALGSP----- - - - - DRTGSVSVIVAVSPAGGDFSDPVTATLGIVQVFWGLDKKLAQRK
 vvc_Hvu LASFYERAGRVKQCLGSP----- - - - - DRTGSVTVIGAVSPPGDFSDPVTSAFLSIVQVFWGLDKKLAQRK
 vvc_Zma LASFYERAGRVKCLGSP----- - - - - DRNGSVTVIGAVSPPGDFSDPVTSAFLSIVQVFWGLDKKLAQRK
 vvc_Bta LASFYERAGRVKCLGNP----- - - - - EREGSVSVIVGAVSPPGDFSDPVTSAFLGIVQVFWGLDKKLAQRK
 vvc_Dme LATFYERAGRVKCLGNP----- - - - - EREGSVSVIVGAVSPPGDFSDPVTSAFLGIVQVFWGLDKKLAQRK
 vvc_Tco LASFYERAGRVTCLGGP----- - - - - KREGSVTVIGAVSPPGDFSDPVTSAFLGIVQVFWGDKKLAQRK
 avc_Mja LAQFYERAGRVTCLGKD----- - - - - NRQGFNCIVGAVSPPGDFSEPVTSNTLRIVKVFWALDANLARR
 evc_Ehi LAEYYERSGRVIALGSD----- - - - - QREGSITAISAVSPSGGDISEPVTQNTLRIVKVFWALDAKLSQRR
 avc_Mba LAEYYERAGYAESLCGE----- - - - - TGSITVIGAVSPPGDFSEPVQTQNTLRIVKVFWALDAKLSQRR
 avc_Mma LAEYYERAGVAETLCGE----- - - - - KGSITIAIGAVSPPGDFSEPVNTQNTLRIVKVFWALDAKLSQRR
 avc_Hsa LSEFYERAGYFENFNGT----- - - - - EGSISIVGAVSPPGDFSEPVQTQNTLRIVKTFWALDSQAFRR
 avc_Hvo LSEFYERAGYFTTVNGE----- - - - - EGSISIVGAVSPPGDFSEPVQTQNTLRIVKTFWALDAQAFRR
 evc_Tth LAAFYERAGRKVITLGG----- - - - - EGAVTIVGAVSPPGDFMSEPVQTQNTLRIYGAFWRLDASLAQAFRR
 avc_Sac LAEYYERAGRVIDGNP----- - - - - ERYGSVIAASAVSPPGDFTEPVTSNTLRFVRVFWPLDVSLAQAR

vvr_Ncr I PPPINVLPPLSFLMKS AIG--EGMTRKDHGDSNQLYAKYAIGRDAAMKAVVCEEALSNE DKL SLEFL
 vvr_Sce I YPPINVLPPLSRLMKS AIG--EGMTRKDHGDSNQLYAKYAIGNDAAAMKAVVGEELSIEDKL SLEFL
 vvr_Bta I YPPINVLPPLSRLMKS AIG--EGMTRKDHSVDVSNQLYACYAIGNDQAMKAVVGEELTSDDLLYLEFL
 vvr_Dme I YPPVNVLPSLSRLMKS AIG--EGMTRKDHSVDVSNQLYACYAIGNDQAMKAVVGEELTPDDLLYLEFL
 vvr_Hvu I YPPINVLPPLSRLMKS AIG--EGMTRRDHSVDVSNQLYANYAIGNDQAMKAVVGEELSS EDLLYLEFL
 vvr_Tco L YPPINILPPLSRLMKN AIG--EGMTRRDHSGSNQMYARMRLD-DILAMKAVVGEETLSS EDLLHLEFL
 avr_Mba I YPPINVLPPLSRLMNSGIG--AGKTREDHKAVSDQMYAGYAEGRDLRGLVAIVGKEALSERDVKFLEFA
 avr_Mma I YPPINVLPPLSRLMNSGIG--AGKTREDHKAVSDQMYAGYAEGRDLRGLVAIVGKEALSERDTKFLEFA
 avr_Hsa VTPPVNVLPPLSRLMDDGIG--EGLTRADHGDSQDLYAAAYAEGEELRDLVNIVVGREALSERDNRYLDFA
 avr_Hvo T RPPINPLPPLSRLMDDGIG--EGLTREDHADVSQDMHYAAAYAEGEDLRDLVNIVVGREALSERDNKYLDFA
 avr_Mja I YPPIVDVLPPLSRLLAGNQG--PGKTREDHKKVVNQAYAAQGRSLRDLVAVVGEELTDRDRAYLKFA
 evr_Ehi I QPPIDPLPPLSRLKDKGTG--AGKTREDHAATMNQLFAAYAQGKQAKELAVVLGESALSDIDKIYAKFA
 evr_Tth I YPPIDPLPPLSRLMNNGVG--KGKTRDHDQVSDQLYSAYANGVDIRKLVATIGEDALTENDRRYLQFA
 avr_Sac I YPPINVLPMSRLMKD GIG--EGKTRDDHKDLNSQNLFAAYARAQDIRGLAAIIGEDSLSEVDRKYLFA
 cfr_Osa I RPAINVGI SVSRVGSAAQI--KAM----KQVAGKSKLELAQFAELQAFQFASA--LDKTSQNQLARG
 cfr_Zma I RPAINVGI SVSRVGSAAQI--KAM----KQVAGKSKLELAQFAELQAFQFASA--LDKTSQNQLARG
 cfr_Tae I RPAINVGI SVSRVGSAAQI--KAM----KQVAGKSKLELAQFAELQAFQFASA--LDKTSQNQLARG
 cfr_Nta I RPAINVGI SVSRVGSAAQI--KAM----KQVAGKSKLELAQFAELQAFQFASA--LDKATQNQLARG
 cfr_Sol I RPAINVGI SVSRVGSAAQI--KAM----KKVAGKLLLELAQFAELLEAFQFASD--LDKATQNQLARG
 cfr_Psa I RPAINVGI SVSRVGSAAQI--KAM----KQVAGKLLLELAQFAELLEAFQFASD--LDKATQNQLARG
 cfr_Mpo I RPAINVGI SVSRVGSAAQI--KAM----KQVAGKLLLELAQFAELLEAFQFASD--LDKATQNQLARG
 cfr_Cpa L RPAINVGI SVSRVGSAAQI--KAM----KQVAGKLLLELAQFDELKA FQS FSSD--LDKATQLQLAG
 cfr_Ppu I RPAINVGI SVSRVGSAAQI--KAM----KQVAGKLLLELAQFAELLEAFQFASD--LDKATQNQLARG
 efr_Syn L RPAINAGI SVSPVGSAQI--KAM----KQVAGKLLLELAQFAELLEAFQFASD--LDKATQNQLARG
 efr_Bsu VRPA INAGI SVSRVGGSAQI--KAM----KKVSGTLRLD LASYRELEAFQFGS D--LDQATQAKLNPG
 efr_Ehi TRPAVDAGL SVSRVGGSAQI--KAM----KKVAGT LRLD LASYRELEAFQFGS D--LDAATQAKLNPG
 efr_Pmo FRPA INAGI SVSRVGGSAQI--KAM----KQVAGT LRLD LASYRELEAFQFGS D--LDAATQAKLNPG
 mfr_Osa IRPA INVGL SVSRVGSAAQL--KAM----KQVAAQVKLELAQYNE LLTFAQFGS D--UDKATLAQLERG
 mfr_Zma IRPA INVGL SVSRVGSAAQL--KAM----KQVC GSSKLELAQYREVA AAQFQFGS D--LDAATQALLNRG
 mfr_Bta IRPA INVGL SVSRVGSAAQT--RAM----KQVAGTMKLELAQYREVA AAQFQFGS D--LDAATQQLLSRG
 mfr_Hsa IRPA INVGL SVSPVGSAQT--RAM----KQVAGTMKLELAQYREVA AAQFQFGS D--LDAATQQLLSRG
 mfr_Dme IRPA INVGL SVSRVGSAAQT--KAM----KQVAGSMKLELAQYREVA AAQFQFGS D--LDAATQQLLN RG
 mfr_Spu I PPAP INVGL SVSRVGSAAQV--KAM----KQVAGSMKLDLGEYREVA AAQFQFGS D--LDASTQNLLNRG
 mfr_Ncr VRPA INVGL SVSRVGSAAQL--KAM----KQVAGSLKLF LAQYREVA AAQFQFGS D--LDAATRQTLNRG
 efr_Rbl IRPA VNTGL SVSRVGSAAQT--SAM----KSVAGKVKLELAQYREMAAAFAQFQFGS D--LDASTQALLNRG
 mfr_Cre I RPA LN VGL SVSRVGSAAQF--PGM----KQVAGT LKLELAQYREVA AAQFQFGS D--LDAATQYVLERG
 efr_Sli QRPA LN VGI SVSRVGGSAQH--KAM----KQVSGRLRVDLAQFRELEAF AAFQFGS D--LDAASNSQLERG
 efr_Eco I RPAP VNP G I SVSRVGGAAQT--KIM--KKLSSGGIR TALAQYRELAFAFSQFASD--LDDATRNQLDHG
 afr_Mba VLPAP DVG KSVSRVGGKAQL--AAY----REVAGDLKLAYS QFEELETFRGAK--LDENTRRLIEHG
 cfc_Osa I YPAVDPLD STSTM LQP----RIVGNEHYETAQRVKQTLQRYKELQDI IAI GLDELSEEDRLTVARA
 cfc_Zma I YPAVDPLD STSTM LQP----RIVGNEHYETAQRVKETLQRYKELQDI IAI GLDELSEEDRLTVARA
 cfc_Tae I YPAVDPLD STSTM LQP----RIVGNEHYETAQRVKETLQRYKELQDI IAI GLDELSEEDRLTVARA
 cfc_Nta I YPAVDPLD STSTM LQP----RIVGEEHYETAQRVKETLQRYKELQDI IAI GLDELSEEDRLTVARA
 cfc_Sol I YPAVDPLD STSTM LQP----RIVGEEHYETAQRVKETLQRYKELQDI IAI GLDELSEEDRLTVARA
 cfc_Psa I YPAVDPLD STSTM LQP----RIVGEEHYETAQRVKETLQRYKELQDI IAI GLDELSEEDRLTVARA
 cfc_Mpo I YPAVDPLD STSTM LQP----WIVGEEHYETAQGVKQTLQRYKELQDI IAI GLDELSEEDRLTVARA
 cfc_Pth I YPAVDPLD STSTM LQP----WIVGEEHYETAQGVKQTLQRYKELQDI IAI PGLDELSEEDRLTVARA
 cfc_Cpa I YPAVDPLD STSTM LQP----GIVGEKH YACAQRVKGILQRYKELQDI IAI SILQ LDELSEEDDRLLAVARA
 cfc_Ppu I YPAVDPLD STSTM LQP----GIVGTDHYSTAQEVKSTLQRYKELQDI IAI GLDELSEEDRQTVSRA
 efc_Syn I YPAVDPLGSTSNMLQP----DIVGSEHYQTARAVQATLQRYKELQDI IAI GLDELSEEDRLTVARA
 mfc_Osa I YPAVDPLDSTS RMLSP----HV LGEDHYNTARGVQ RVLQNYKNLQDI IAI LGMD ELS EDDRLTVRA
 mfc_Zma I YPAVDPLDSTS RMLSP----HV LGEDHYNTARGVQ KV LQNYKNLQDI IAI LGMD ELS EDDKLTVARA
 mfc_Nta I YPAVDPLDSTS RMLSP----HV LGEDHYNTARGVQ KV LQNYKNLQDI IAI LGMD ELS EDDKLTVARA
 mfc_Hbr I YPAVDPLDSTS RMLSP----HV LGEEHYNTARGVQ KV LQNYKNLQDI IAI LGMD ELS EDDKLTVARA
 mfc_Bta I YPAVDPLDSTS RIMDP----NIVGSEHYDVARGVQ KILQDYKSLQDI IAI LGMD ELS EDDKLTVARA
 mfc_Rno I YPAVDPLDSTS RIMDP----NIVGSEHYDVARGVQ KILQDYKSLQDI IAI LGMD ELS EDDKLTVSRA
 mfc_Sce I YPAVDPLDSTS RMLD----AVVGQEHYDVASKVQ ETLQTYKSLQDI IAI LGMD ELS SEQDKLTVERA
 mfc_Rbl I YPAVDPLDSTS RMLD----QILGEEHYNTARAVQGQ ILQRYKNSLQDI IAI LGMD ELS SEEDKLTVARA
 mfc_Ncr I YPAVDPLDSTS RMLD----RIVGQEHYETATRVQQILQ EYKSLQDI IAI LGMD ELS SEADKLTVARA
 efc_Eco I YPAVDPLDSTS RQLD----LUVGQEHYDTARGVQ S I LQRYQ EKLQDI IAI LGMD ELS SEEDKLTVVVARA
 efc_Pmo I YPAVDPLDSTS TA LE P----QITGEHYNTAREVQ QILQRYKELQDI IAI LGMD ELS SEEDKNTVNR
 efc_Bsu I YPAVDPLASTS RALAP----EIVGEEHYAVAREVOSTLQRYKELQDI IAI LGMD ELS GEEDKLVVHRA
 efc_Ehi I YPAVDPLASS S S A L P----EIVGEEHYKATEVQ H V L QRYRELQDI IAI LGMD ELS SDQE KLV V SRA
 efc_Sli I YPAVDPLDSTS RIL D P----RYIAAEHYNAAMRVK N I LQKY K D L QDI IAI LGMD ELS GEEDKLVVHRA
 afc_Mba L YPAIDPLQNSK M A T P----GIVGERHYLLAQEIROT LAQYSELKDI ISMLGLEQLSPEDRN V V R A
 vvc_Ncr HFP SINTSVSY SKYLTILDWKWY-E REY P D F P RLDRIRQLLSDSEEDLQV V Q L YG K S A L S D P D N I T L D M A
 vvc_Sce HFP SINTSVSY SKYTNVLNKFY-DSNYP EFPV LDR DM K E ILSN AEE E L Q V V Q L V G K S A L S D S D K I T L D V A
 vvc_Hvu HFP S V N W L I S Y SKY STALE GYY-EKFDPG FID M RT K A R E V L Q R E D D L N E I V Q L V G K D A L G E S D N I T L E T A
 vvc_Zma HFP S V N W L I S Y SKY SKA L E SFY-EKFDPDFID IRT K A R E V L Q R E D D L N E I V Q L V G K D A L A E S D K I T L E T A
 vvc_Bta HFP S V N W L I S Y SKY M R A L D E YY-DKHFTEF VPLRTK A E I L Q E E E D L A E I V Q L V G K A S L A E T D K I T L E V A
 vvc_Dme HFP S V N W L I S Y SKY M R A L D E YY-DK NY P E F V P L R T K V K E I L Q E E E D L S E I V Q L V G K A S L A E T D K V T L E V A
 vvc_Tco HFP S V N W L I S Y SKY L N A L E P F F -N T L D P D Y M R L R S V A A E I L Q R E E E L Q E I V Q L V G K D S L S E S D K I T L E T A
 avc_Mja HFP A I N W L Q S Y SLY I D D V E W W N T N T G P D W R Q L R D E A M S L L Q K E A E L Q E I V Q L V G P D A L P D R E R V I L E V A
 evc_Ehi HFP A I N W I Q S Y SLY S T E V G R Y M D Q I L Q Q D W S D M V T E G M R I L Q E E E Q L N E I V R L V G I D S L S D N D R L T L E V A
 avc_Mba HFP A I N W L N S Y SLY K D S L N D W F A D N V A P D Y Y P L R E A M E L Q T E S E L Q E I V Q L V G S D A L P D D Q Q L L L E I T
 avc_Mma HFP A I N W D E S Y SLY K D Q L D P W F T D N V V D D W A E Q R Q S A V D I L D E E S E L E E I V Q L V G K D A L P E D Q Q L T L E V A
 avc_Hsa HFP A I N W N E S Y SLY Q E Q L D P W F V E N E D D W A E E R Q W A V D L D E E N E L Q E I V Q L V G K D A L P E D Q Q L T L E V A
 avc_Hvo HFP A I N W H Q S Y L F T S A L D P W Y R E N V A E D Y P E L R O A I S E L L Q E I V Q L V G P D A L Q D A E R L V I E V G
 evc_Tth HFP A I N W I Q G F S A Y V D L V A Q W W H K N V D P N W K E M R D T M M K V L I R E D E L R Q I V R L V G P F E L A E K D K L V I E A A

vvr_Ncr + SFIAGQGPYESRTIFESLDLAW + SLLRIYRKDMLN RIPKKIIDEFYSRSAADRK
 vvr_Sce EKFEK-----TFITQGAYEDRTVFEISLDQAW---SLLRIYPKEMLN RISP KILDEFYDR-ARDDA
 vvr_Bta QKFEK-----NFITQGPYENRTVYETLDIGW---QLLRIFPKEMLKRIPIQSTLSEFYPRDSAKH-
 vvr_Dme TKFEK-----NFISQGNYENRTVFEISLDIGW---QLLRIFPKEMLKRIPIASILADEFYPRDSRH--
 vvr_Hvu DKFER-----KFVAQGAYDTRNIFQSLDLAW---TLLRIFPRELLHRIPAKTL DQFYSRDATH--
 vvr_Tco EKFEK-----KFIGQGFHESRDVFQSLDLCW---QLLRIFPKVLLNKIDVXTRNEFYDGQPMRC-
 avr_Mba DLFEQ-----QFV1QGRNENRTIA DTL DIGW---KILAHLPE NQLGRIDNKYI QKYHPAHRKGQ-
 avr_Mma DLFED-----K FVRQGRNENRTIA DTL DIGW---QILTHLPE NQLGRIDNKYI QKYHPAHRKAK-
 avr_Hsa DRFEA-----EFIGDQGFKTNRDIEETL DLGW---ELLSMFPKTELNR DEDLIEDHYVEDYADEA
 avr_Hvo ERFEA-----EFVNQGFD TDRSIEETL DLGW---DLLSTLPKSELNR IDEELIEDDY-Y-EDDAESV
 avr_Mja DEFED-----KFVRQGNDEDRSIEETL DLLW---ELLAILPEEELKRVRELIEKYHPN YRKK-
 evr_Ehi ERFEN-----EYNQGFTYNTRTITEETL DLGW---ELLALMPRT EKRIKDOLLDKYLPEGK-
 evr_Tth DAFER-----FFINQG-QQNRSIEESLCIAW---ALLSMLPQGEKLRSKDHIGKYYGQKLE EIW
 avr_Sac EA FER-----RFVAQGVVNENRSIETL DLGW---EVLSILPESEL SLIRSEYIKKYHPN YRGKK-
 cfr_Osa RRL-R ELLKQSQANPLPVEEQIATIYIGTRGYLDSLEIGQVKKFDEL RKHLLKDTKPQFQE
 cfr_Zma RRL-R ELLKQSQSNPLPVEEQIATIYIGTRGYLDSLEIEQVKKFDEL RKHLLKDTKPQFQE
 cfr_Tae RRL-R ELLKQSQANPLPVEEQIATIYIGTRGYLDSLEIEQVKKFDEL RKHLLKDTKPQFQE
 cfr_Nta QRL-R ELLKQSQSA PLTVEEQIATIYIGTRGYLDSLEIEQVKKFDEL RKHLLKDTKPQFQE
 cfr_Sol QRL-R ELLKQPQSA PLTVEEQIATIYIGTRGYLDSLEIEQVKKFDEL RKHLLKDTKPQFQE
 cfr_Psa QRL-R ELLKQSQSA PLTVEEQIATIYIGTRGYLDSLEIEQVKKFDEL RKHLLKDTKPQFQE
 cfr_Mpo QRL-R ELLKQSQSA PLSVEEQIATIYIGTRGYLDSLEIEQVKKFDEL RKHLLKDTKPQFQE
 cfr_Cpa ERL-R ELLKQQYAPLPVVEEQIATIYIGTRGYLDSLEIEQVKKFDEL RKHLLKDTKPQFQE
 cfr_Ppu QRL-R EILKQAQNSP IPVVEEQIATIYIGTRGYLDSLEIEQVKKFDEL RKHLLKDTKPQFQE
 efr_Syn QRL-R EILKQPQYSP IPVVEEQIATIYIGTRGYLDSLEIEQVKKFDEL RKHLLKDTKPQFQE
 efr_Bsu ART-V EVLKQDLNKP LPVEKQV AILYALT KGYLDDIPVADIRR FE EYYMYLDQNHNDL
 efr_Ehi RRT-V EILKQKLHAPLA VEKQV VILYALT HFLDSIPVDSILD FEH ELYDTN HADIFE
 efr_Pmo HRI-M EILKQEQYKPFVVEEQVVS FTVINGYLD DIAIDQV RRF E-----KELLE
 mfr_Osa ARL-T EVPKQPQYEP LPIEKQIVVIYAAVNGFCDRMLDRISQYEKAILSTINPELLKSFN
 mfr_Zma ARL-T EVPKQPQYEP LPIEKQIVVIYAAVNGFCDRMLDRISQYEKNILSTINPELLKSFL
 mfr_Bta VRL-T ELLKQGQYSPMAIEEQVAVIYAGVRGYLOKLEPSKITKFEN AFLSHVISQHQALLS
 mfr_Hsa VRL-T ELLKQGQYSPMAIEEQVAVIYAGVRGYL D KLEPSKITKFEN AFLSHVISQHQALLG
 mfr_Dme VRL-T ELLKQGQYVPM AIEEQVAVIYAGVRGYL D KMDPAK ITKF EFLQH IKTSEQ ALLD
 mfr_Spu VRL-T ELLKQGQYTPM AIEEQVAVIYAGVRGYL D KMDPAK ITKF EFLQH IKTSEQ ALLD
 mfr_Ncr ERL-T ELLKQKQYSPMAV NEMVPLIFAGVNGFLDSV PAKILQWEADFLAH LKTNEPEIMA
 efr_Rbl APL-T ELMKQPQYSP LTNAEIVCII FAGTNGYLDKVAV RDVG RFEAG LKH LR TNGN ALLD
 mfr_Cre ARL-T EVLKQKQFVPMPIEQQTIVVYATKGYLDKVAV RDVG RFEAG LKH LR TNGN ALLD
 efr_Sli QRM-V ELLKQNQYPM STEDQVSVWAGTGRNM D E VPYADIRR FE KELLEY LHR QEQ GLMT
 efr_Eco QKV-T ELLKQKQYAPMSVAQSQSLVFAAERGYLADVEL SKIGS FEA ALLAY VDRDHA PLMQ
 afr_Mba RRI-R ALLKQPHNSPVPVDPQIVI LVALNAKLFDNVPLDKMVEAE NSLRKALPDIPEDVRE
 cfc_Osa RKİERFLSOPFFVVAEVFTGSPGKYVGLAETIRGFQ LILSGE LDGLPEQAFYLVGN-IDEASTKAINLEE
 cfc_Zma RKİERFLSOPFFVVAEVFTGSPGKYVGLAETIRGFQ LILSGE LDGLPEQAFYLVGN-IDEASTKAINLEE
 cfc_Tae RKİERFLSOPFFVVAEVFTGSPGKYVGLAETIRGFQ LILSGE LDGLPEQAFYLVGN-IDEASTKAITLEE
 cfc_Nta RKİERFLSOPFFVVAEVFTGSPGKYVGLAETIRGFQ LILSGE LDGLPEQAFYLVGN-IDEATAKAMNLEM
 cfc_Sol RKİERFLSOPFFVVAEVFTGSPGKYVGLAETIRGFQ LILSGE LDGLPEQAFYLVGN-IDEATAKAMNLEM
 cfc_Psa RKİERFFSOPFFVVAEVFTGSPGKYVGLAETIRGFQ LILSGE LDGLPEQAFYLVGN-IDEATAKATNL-
 cfc_Mpo RKİERFLSOPFFVVAEVFTGSPGKYVSL RETIKGFQ MILSGE LDGLPEQAFYLVGN-IDEATAKAATLQV
 cfc_Pth RKİERFLSOPFFVVAEVFTGSPGKYVGL METIRGFQ MILSGE LDGLPEQAFYLVGN-IDEATAKAINSM
 cfc_Cpa RVERFLSOPFFVVAEVFTGSPGKYVGSLED TIKGFTMILDGE LDGLPEQAFYLVGD-IOEAISNGQKLLA
 cfc_Ppu RKİERFLSOPFFVVAEVFTGSPGKYVSL EDAIKGFQ MILKGEL DDLPEQAFYLVGD-IDEAIQKADSMKD
 efc_Syn RVERFLSOPFFVVAEVFTGAPGKYV TLEETIKGFQ MILSGE LDGLPEQAFYMVGN-IEFAKA EAKLKA
 mfc_Osa RKİQRFLSOPFHVVAEVFTGAPGKYV ELKESVNSFQGVLD GKYDDLPEQSFY MVGG-IEEVIAKA EKI AK
 mfc_Zma RKİQRFLSOPFHVVAEVFTGAPGKYV ELKESVNSFQGVLD GKYDDLPEQSFY MVGG-IEEVIAKA EKI AK
 mfc_Nta RKİQRFLSOPFHVVAEVFTGAPGKYV ELKESVNSFQGVLD GKYDDLSEQSFY MVGG-IDEVIKA EKI AK
 mfc_Hbr RKİQRFLSOPFHVVAEVFTGAPGKYV ELKESITSFQGVLD GKYDDLPEQSFY MVGG-IDEVIKA EKI AK
 mfc_Bta RKİQRFLSOPFQVAEVFTGHL GKL VPLKETIKGFQ QI LAGEYDHLPEQAFY MVGP-IEBAVAKADK LAE
 mfc_Rno RKİQRFLSOPFQVAEVFTGHMGKL VPLKETIKGFQ QI LAGEYDHLPEQAFY MVGP-IEBAVAKADK LAE
 mfc_Sce RKİQRFLSOPFQVAEVFTGIPGKL VRLK DTV SFKA VLE GKYD NIP E HAFY MVGG-IEDVVAKAEK LAA
 efc_Rbl RKİQRFLSOPFQDV A QVFTGSPGVQVPL DKTIASFKAV VAGEYDHLPEA AFY MVGD-IEBAKAKAAK LAA
 mfc_Ncr RKİQRFLSOPFTVAQVFTGIEGKL VDLK DKTIA SFNAIL AGE GD DLP EGA FY MVGD-FASARAKGEK ILA
 efc_Eco RKİQRFLSOPFQVAEVFTGSPGKYVSLK D TIRGFNGIM EYDHLPEQAFY MVGS-IEBAVAKAKKL-
 efc_Pmo RKİERFFSOPFHVVAE QFTGMDGKYVTVKETIRGFKE II EGRH DDLPEQAFYLVGT-IDEBAIAKARELMK
 efc_Bsu RRIQFFLSQNFNVAE QFTGOKGSYVPKETVQGFK EILAG KYD H DLP EAFRLVGR-IEEVVEKA KENG V
 efc_Ehi RRVQFFLSQNFNVAE QFTGOKGSYVPKETVQGFK EILAG KYD H DLP EAFRLVGR-IEEVVEKA KENG V
 efc_Sli RRVERFLSOPFTVAKQFTGVDGSDVPL D E SIA FN AIC D GYDH FPEQAFY MC GG-IE DLKNNAKELG W
 afc_Mba RRLERFLTOPFTTEQFTGINKGSISLSDALG CERILN D E FDYD LYMIGT-IDEAMAKKSSREK
 vvc_Ncr T-LIKE DFLQQNGYSDY-DQFCPIWKTEWMMLMMG-FHDEAQKAIAGQGN-----WNK VRE
 vvc_Sce T-LIKE DFLQQNGYSTY-DAFCPIWKTFDMMRAFIS-YHDEAQKAVANGAN-----WSKLAD
 vvc_Hvu K-LREDYLAQNAFTPY-DKTCPPYKICWMLRNIV A-FYDEAQKAVANGAN-----WSKLAD
 vvc_Zma K-LREDYLAQNAFTPY-DKFCPFYKCSWM MRNIIH-FNOLANQAVERAAN-ADGHKITYAVVKS
 vvc_Bta K-LIKDFLQQNGYTPY-DRFCPFYKTVGMLNSMIA-FYDMARRA VETTAQ-SD-NKITWNSIIRE
 vvc_Dme K-LLKDFLQQNSYSPY-DRVC PFYKTVGMLRNIV A-FYETARHCLESTAQ-SD-NKITWNTIKE
 vvc_Tco K-VIREEFLQQNAFTPY-DKTCPPYKICWMLRNIV A-FYEE SQRVVAESA-GELKITWNYIRE
 avc_Mja R-MLREDFLQQNAFDEV-DTYPMPKQYMLK IIMT-FYQEALKAV ERGVE-PAKILG-VSVKQ
 evc_Ehi K-SIREDFLQQNAFDDV-DTFTSREKQFNM LK VILT-FGKEARKALSLGAY-FNEIME GTVAVRE
 avc_Mba R-MLREIFLQQNAFHPV-DAYSPFDQYKILKAIMK-WGDAAMDALKSGVP-VTEIK-LES KN
 avc_Mma R-MIREIFLQQNAFHP -DTYSPF EKQYKIMK-AIMK-WGDAAMDALKSGVP-SSEILK-IES KD
 avc_Hsa P-YIREAWLQQNALHDV-DRCPPEKTYA ILSGIKT-LHEESFEAL DAGV GP-VEEITS-IDA AP
 avc_Hvo R-YIREAYLQQNAFHP -DTYCSPEKTYA ILSGIKT-FNDEAFKALEAGV GP-VEEIQ A-IEAAP
 evc_Tth R-IREDFLQQNAYHEV-DAYCSMKKAYGIMKMLA-FYKEAEAAIKRGV S-IDEILQ-LPVLE
 avc_Sac K-LIKDAFLKENAYDDI-DAFSSPQKQVRIMRLIYI-FYNQS QD LISKGV P-LKKILD VGP IEP

vvv_Ncr	GKGKD KPTTKDTRDTAAPEEEENLIDA
vvv_Sce	D EDEEDPTQEAPVRRRTPAKKNL
vvv_Bta	-
vvv_Dme	-
vvv_Hvu	-
vvv_Tco	-
avr_Mba	-
avr_Mma	-
avr_Hsa	TAD-
avr_Hvo	E AER
avr_Mja	-
evr_Ehi	-
evr_Tth	GAPQALD-
avrSac	-
cfr_Osa	I I SS-SKTFTEEEAEILLKEAIQEQLERFSLQEQT
cfr_Zma	I I SS-SKTFTEQAETLLKEAIQEQLERFSLQEQT
cfr_Tae	I I SS-SKTFTEQAIIILKEAIQEQLERFSLQ
cfr_Nta	I I SS-TKTFTEEEAEALLKEAIQEQMDRFLQEQQA
cfr_Sol	I I SS-TKTFTEEEAEALLKEAIQEQMERFLLQEQQA
cfr_Psa	I I SS-TKTFGEAEAILKEAIQEQMELF--
cfr_Mpo	IIRS-TKVFTEQAEENLLKEAITEHIELFLFQEEK
cfr_Cpa	IIRS-EKALTAEAENLLKDAISDCKQAFLSNI
cfr_Ppu	SIRS-SKKLDTASEELLKKAIEDVKQGFVKA
efr_Syn	A RS-SQKLDETAEALLKEAIAEYKAGFTA-
efr_Bsu	GIAK-TGNLPADED--FKAIAIEGFKRTFAPS
efr_Ehi	TIRT-TKDLPEEE--LNQAIQEKDIFLATKGNTSSTEDKLKSIQNA-
efr_Pmo	E LKT-TQLFS-
mfr_Osa	EKGGLTNERKIEPDASL KQTAKINE
mfr_Zma	EKGGLTNERKMEPDASL KESALNL
mfr_Bta	KIRTDG-KISEESDAKL KIEIVTNFLAGFEA
mfr_Hsa	TIRADG-KISEQSDAKL KIEIVTNFLAGFEA
mfr_Dme	TIAKDG-AISEASDAKL KDIVAKFMSTFQG
mfr_Spu	TIRTEG-QLSPQTEATLKDVVLKF LDTFEG
mfr_Ncr	TIDKEG-AISKD LEAKLRDVIQTFTKSFLG
efr_Rbl	DMTKNDRKVAGD LEKAIRAELDAFAKDFA
mfr_Cre	AKGKITPEINAHLAQ QQMSNLPVMTK
efr_Sli	SIREGGK-MSDDTL QAAEAIAAFKKKQFETSDGKLLGEDAPSAAK
efr_Eco	F I-NQTGGYND EIEGKLGILD SFKATQS
afr_Mba	RFKG-D KELSDKDRE TILNIARKALEPYQPKESESKPEAKTEENAESETQTEDKSESEAKTEEVAKSE
cfc_Osa	E NKLKK--
cfc_Zma	E SKLKK--
cfc_Tae	E NKSQK--
cfc_Nta	E SNLKK--
cfc_Sol	E SKLKK--
cfc_Psa	--
cfc_Mpo	E S--
cfc_Pth	E S--
cfc_Cpa	E AK--
cfc_Ppu	--
efc_Syn	--
mfc_Osa	E SAS--
mfc_Zma	E SAS--
mfc_Nta	E SA--
mfc_Hbr	E SAS--
mfc_Bta	E HS--
mfc_Rno	E HGS--
mfc_Sce	E AN--
efc_Rbl	A AA--
mfc_Ncr	E LEGQA--
efc_Eco	--
efc_Pmo	G AE--
efc_Bsu	E V--
efc_Ehi	--
efc_Sli	S --
afc_Mba	S --
vvv_Ncr	ATQDLQAQLKSLKF EVPS-EGQE KICKKYEA IQQQM LDKFASVIDE
vvv_Sce	STGDVKHAVSSSKFFEPS-R GEKEVHGE FEKLLSTM QERFAESTD -
vvv_Hvu	RGDLFYRLVSQKF EDPA-EGEDV LVKFQKLYDDLTAGFRNLEDEAR
vvv_Zma	RLGDLFYRLVSQKF EDPA-EGEEAL VGKFKKLYDDLTGFRNLEDEAR
vvv_Bta	HMGEI LYKLSSMKFKDPVKDGEAKIKADY YA QLEDQMNAFRSLED -
vvv_Dme	SMGGIMYQLSSMKFKDPVKD GEQKIKADYDQLYEDLQQAFRNLED -
vvv_Tco	M I PHIYTGLTEMKFRDP-Q E GEEANVEFYRKQNEEIVSAFASLLQ-
avc_Mja	DIARMKY-IP HDEFINVKS -KEIMEKIKNELGSLN--
evc_Ehi	RISRSKY- I P <small>E</small> EEELAKISSINEE----IKETIQLIVSEGGM TDD -
avc_Mba	VLA KV--KYEEKF ----DES---MNAVLAQMD KEFASLRGR -
avc_Mma	SLPKV---KYEENF----EDS---LNAVLA KMDKEFAALGGK -
avc_Hsa	RLNR LGT-TPDDEH ----EAE----VAEIKQQITEQLRELY--
avc_Hvo	RLNRIGV---QEDW----EAY----IEDL KAEITEQLRELY --
evc_Tth	RIGRARY-VSEEEF----PAY----FEEAM KEIQGAFKAGLGERWTF --
avc_Sac	E IRIKYTIKND E LNKIDEIENK---LKATFD SLLKEVS --

	+	+	+
vvr_Ncr	-	-	-
vvr_Sce	-	-	-
vvr_Bta	-	-	-
vvr_Dme	-	-	-
vvr_Hvu	-	-	-
vvr_Tco	-	-	-
avr_Mba	-	-	-
avr_Mma	-	-	-
avr_Hsa	-	-	-
avr_Hvo	-	-	-
avr_Mja	-	-	-
evr_Ehi	-	-	-
evr_Tth	-	-	-
avr_Sac	-	-	-
cfr_Osa	-	-	-
cfr_Zma	-	-	-
cfr_Tae	-	-	-
cfr_Nta	-	-	-
cfr_Sol	-	-	-
cfr_Psa	-	-	-
cfr_Mpo	-	-	-
cfr_Cpa	-	-	-
cfr_Ppu	-	-	-
efr_Syn	-	-	-
efr_Bsu	-	-	-
efr_Ehi	-	-	-
efr_Pmo	-	-	-
mfr_Osa	-	-	-
mfr_Zma	-	-	-
mfr_Bta	-	-	-
mfr_Hsa	-	-	-
mfr_Dme	-	-	-
mfr_Spu	-	-	-
mfr_Ncr	-	-	-
efr_Rbl	-	-	-
mfr_Cre	-	-	-
efr_Sli	-	-	-
efr_Eco	-	-	-
afr_Mba	T Q T K G K P E T E T K T E E K P E F G T R T E E K	-	-
cfc_Osa	-	-	-
cfc_Zma	-	-	-
cfc_Tae	-	-	-
cfc_Nta	-	-	-
cfc_Sol	-	-	-
cfc_Psa	-	-	-
cfc_Mpo	-	-	-
cfc_Pth	-	-	-
cfc_Cpa	-	-	-
cfc_Ppu	-	-	-
efc_Syn	-	-	-
mfc_Osa	-	-	-
mfc_Zma	-	-	-
mfc_Nta	-	-	-
mfc_Hbr	-	-	-
mfc_Bta	-	-	-
mfc_Rno	-	-	-
mfc_Sce	-	-	-
efc_Rbl	-	-	-
mfc_Ncr	-	-	-
efc_Eco	-	-	-
efc_Pmo	-	-	-
efc_Bsu	-	-	-
efc_Ehi	-	-	-
efc_Sli	-	-	-
afc_Mba	-	-	-
vvc_Ncr	-	-	-
vvc_Sce	-	-	-
vvc_Hvu	-	-	-
vvc_Zma	-	-	-
vvc_Bta	-	-	-
vvc_Dme	-	-	-
vvc_Tco	-	-	-
avc_Mja	-	-	-
evc_Ehi	-	-	-
avc_Mba	-	-	-
avc_Mma	-	-	-
avc_Hsa	-	-	-
avc_Hvo	-	-	-
evc_Tth	-	-	-
avc_Sac	-	-	-

謝 辞

本研究をまとめるにあたり、標品の供与や技術的指導、研究上の協力において多くの研究者の方々のお世話になりました。その中でも特に、取り扱いに特殊な機器や技術を要求する絶対嫌気性のメタン産生菌の培養に関しては、産業医科大学教授・古賀洋介先生の直接のご指導を、また、最終章の系統樹の作成については、統計数理研究所教授・長谷川政美先生のご助言をいただくことができました。本研究を通じて、このような分野ごとの権威の先生がたと関わりを持てたことをたいへんな誇りと思い、今後の研究活動に糧にしてゆきたいと考えております。そして長年にわたり、このような恵まれた機会と研究環境をお与えくださった吉田賢右先生、大島泰郎先生に心からお礼をもうしあげます。