

Title	両端末cDNAライブラリーの作製とその解析
Author(s)	的場, 亮
Citation	大阪大学, 1993, 博士論文
Version Type	VoR
URL	https://doi.org/10.11501/3065776
rights	
Note	

Osaka University Knowledge Archive : OUKA

<https://ir.library.osaka-u.ac.jp/>

Osaka University

①

両末端 cDNA ライブラリーの作製とその解析

細胞生体工学センター 遺伝子情報システム研究部門 的場 亮

両末端 cDNA ライブラリーの作製とその解析

細胞生体工学センター 遺伝子情報システム研究部門 的場 亮

目次

[はじめに]	1 - 2 頁
[両末端 cDNA ライブラリーの特徴]	3 頁
[方法]	4 - 6 頁
Cytoplasmic mRNA の調製	
全長 cDNA ライブラリーの作製	
両末端 cDNA ライブラリーの作製	
PCR(Polymerase Chain Reaction) ダイレクトシーケンス法	
データ解析	
[結果]	7 - 14 頁
両末端 cDNA のサイズ	
シーケンスエラー	
Expression profile	
両末端 cDNA ライブラリーの評価	
3'-directed cDNA ライブラリーとの Similarity	
Protein Database PIR との Homology	
ORF 検索	
Hydropathy 解析	
モチーフ検索	
[考察]	15 - 18 頁
[Fig., Table の説明]	19 - 23 頁
[参考文献]	24 - 25 頁

Fig.1 - Fig.8 , Table 1 - Table 4

付録. シーケンスデータ

[はじめに]

ゲノムプロジェクトにおいて cDNA の解析は、遺伝子として機能している部分の解析という意味で重要である。様々な組織・細胞でどのような種類の遺伝子がどのくらいの頻度で発現しているかを調べることにより、その細胞での遺伝子発現レベルでの特徴をつかむことができる。さらに、遺伝子の発現・調節を中心とするゲノム構造解析の進展と対応して、cDNA はゲノム構造解析にとって有用な遺伝子マーカーとしての役割を果たす。そこで大量の cDNA の塩基配列を決定し、特に未知の遺伝子について分類し、できればその機能を予測できることが期待される。最近の技術的進歩によって大量の未知の遺伝子の部分構造を決めることができる時代に入り^{1),2)}、いくつかのグループが cDNA の大量シーケンスとその解析を行っている^{2),3),4)}。我々のグループでは様々な細胞から mRNA を調製し、その 3' 側平均約 270 base を持つ 3'-directed cDNA ライブラリーを作製し解析を行ってきた。これまでに数千クロンの 3'-directed cDNA ライブラリーのシーケンスを決定し解析した結果、mRNA の 3' 側はユニークであるので、遺伝子の分類・同定と発現量の定量が容易にできることがわかった³⁾。しかし、未知のクローンについては、この 3'-directed cDNA は 3'-noncoding region が長いので、アミノ酸としての情報量が少なく、機能的な予想が困難であった。一方、mRNA の 5' 側は様々な機能的情報を含んでいると考えられ、情報量の濃い領域である。よって、未知の cDNA クローンに対する機能的予測ができる。しかし、5' 側領域のみでは、もし gene family などの異なった遺伝子であるがよく似た配列が存在したとき、その分類・同定が混乱することが予想される。そこで mRNA の 5' 側と 3' 側の両方の長所をいかした、しかも、大量かつ迅速にシーケンスが決定できる cDNA ライブラリーはないかと考えた。5' 側と 3' 側の両方の特徴を持つ cDNA の代表は全長 cDNA であるが、全長 cDNA をシーケンスを決定するためには、PCR 法でテンプレートを増幅させる方法が適用できないので、一旦、大腸菌で増やし DNA を調製しなければならぬ。そこでもっと簡単にシーケンスを決定できる、5' 側と 3' 側の両方の特徴を活かした cDNA ライブラリーを作製しその解析を行った。このデータを利用すれば、将来ゲノムの構造解析が進展したときに遺伝子の両端の位置を同定し、各々に固有の情報を入力することができる。

まず、ヒト肝臓癌由来の細胞株 HepG2 細胞⁵⁾から調製した cytoplasmic mRNA を使って全長 cDNA ライブラリーを作製した。全長 cDNA をプラスミドベクターにライゲーションした後すぐにそれを *Mbo*I (GATC) で切り、cDNA

の真ん中の部分を取り除き(このときベクターの部分はメチレーションによって保護されており切れない)、セルフライゲーションして mRNA の 5' 側と 3' 側の両末端を持つ cDNA ライブラリーを作製した(Fig.1 (a), (b))。全長 cDNA ライブラリーの平均の長さは約 2 kbp であった。一方、MboI (GATC) を用いて切った場合、この両末端 cDNA ライブラリーの平均の長さは約 0.6 kbp であり、ベクターのシーケンスをプライマーとして PCR 法で cDNA 部分を増幅することができる。

PCR ダイレクトシーケンス法を用いてこの両末端 cDNA ライブラリーの塩基配列を決定しその解析を行った。

[両末端 cDNA ライブラリーの特徴]

- * 3' 末端の Poly A から平均約 300 bp 上流に *Mbo*I サイト、そこからさらに平均約 300 bp 上流に 5' 末端が来る構造をとる。
- * ベクター部分の適当なシーケンスをプライマーとして PCR ダイレクトシーケンス法を用いて大量かつ迅速な塩基配列決定ができる。
- * 1 回のシーケンス反応で同一の遺伝子の mRNA の 5' 側と 3' 側の両方のシーケンス情報が得られる。
- * この cDNA の 5' 側はメチオニンに始まるシグナルペプチドなどの protein coding シーケンスを含んでいるので、タンパク質データベースと比較することにより、未知のクローンについての機能的な予測ができる。また、3' 側は 3'-directed cDNA ライブラリーのデータと比較することによってそれらの expression profile 等の情報を利用できる。

[方法]

Cytoplasmic mRNA の調製

まず、肝癌由来の細胞株 HepG2 細胞から cytoplasmic mRNA を調製した⁶⁾。培養した HepG2 細胞を集め、RNA extraction buffer [0.14M NaCl, 1.5mM MgCl₂, 10mM Tris HCl(pH 8.6), 0.5% NP-40, 1mM DTT, 1000 unit/ml RNase inhibitor(Pharmacia)] を加え、30 秒間 Vortex をかけた後すぐ、5 分間氷上に置く。遠心操作(12000×g, 90 秒間)により核やミトコンドリアを沈殿させ、上清に Proteinase K を加えてタンパク質を分解した後フェノール抽出を行なった。RNA はイソプロパノール沈殿し、70%エタノールでリンスを行なった。さらにこの RNA を oligo dT セルロースカラムに通して poly A⁺ RNA を集めた。

全長 cDNA ライブラリーの作製

上記の方法で調製した cytoplasmic mRNA から全長 cDNA ライブラリーを SuperScriptTM Plasmid System (BRL) キットを用いて作製した。

mRNA 2μg を使い *NotI* primer-adapter (5'-GACTA GTTCT AGATC GCGAG CGGCC GCCCT₁₅-3')を用いて SuperScript RT (Reverse Transcriptase) で 1st strand を合成した。そのまま同じチューブで ニックトランスレーション法により 2nd strand を合成した⁷⁾。このとき、*E.coli* DNA ligase, *E.coli* DNA polymerase I, *E.coli* RNase H を同時に反応させた。*E.coli* RNase H で mRNA にニックを入れ、RNA をプライマーとして *E.coli* DNA polymerase I で DNA を合成し、*E.coli* DNA ligase でそれらの DNA 断片をつなげた。この方法により、より長い cDNA が得られる。この反応後、T4 DNA polymerase I で 5' 側を flushing し、フェノール抽出、エタノール沈殿を行なった。T4 DNA ligase で *SalI* adapter (5'-TCGA CCCACG CGTCCG-3' / 3'-GGGTGC GCAGGC-5') を cDNA の 5' 側につなげ、フェノール抽出、エタノール沈殿を行なった。さらに、*NotI* で cDNA の 3' 側の primer-adapter を切り、フェノール抽出、エタノール沈殿を行なった。切れた primer-adapter を除くため、Sephacryl S-500 HR クロマトグラフィカラムに通し、長い cDNA を集め、エタノール沈殿を行なった。作製した cDNA (約10ng) を pSPORT 1 ベクターの *NotI-SalI* サイトにライゲーションし、全長 cDNA ライブラリーが完成した。

両末端 cDNA ライブラリーの作製

上記のライゲーション反応液の半分 (10μl) を取り 10× K buffer [2000mM Tris

HCl(pH8.5), 100mM MgCl₂, 10mM DTT, 1000mM KCl] 4μl, MboI 10unit を加え、最終量 40μl で 37°C, 1 時間反応させた。この反応で cDNA は MboI(GATC) で切られるが、ベクター部分はメチレーションによって保護されており MboI で切れない。次に 65°C, 10 分間反応して酵素を失活させたあと、5μl (1/8 量) を取り、5× *E.coli* DNA ligation buffer [100mM Tris HCl(pH7.5), 500mM KCl, 20mM MgCl₂, 200μg/ml of BSA, 20mM DTT, 750μM βNAD, 50mM (NH₄)₂SO₄] 20μl, *E.coli* DNA ligase 20units を加え、最終量 100μl で 16°C, 6 時間反応させた。このようにして、cDNA の 5' 末端から MboI サイトまでと 3' 末端から MboI サイトまでがライゲーションされて、cDNA の真ん中の部分の抜け落ちた、両末端 cDNA ライブラリーが完成した(Fig. 1(a))。

P C R (Polymerase Chain Reaction)ダイレクトシーケンス法

上記の反応液をコンピテントセル *E.coli* DH5 にトランスフォーメーションを行ない⁸⁾、ampicillin (50μg/ml) + 1.5% Ager +LB 培地のプレートで培養した。37°C, 12 時間培養した後、大腸菌のコロニーを拾い ampicillin (50μg/ml) + 125μl LB 培地の入った 96-well プレートで 37°C, 24 時間培養した。遠心操作によって大腸菌を集めそれに 100μl の水を加えよく混ぜた後、100°C, 20 分間オートクレーブを行った。遠心操作の後、上清を 5μl 取り PCR 反応に用いた。

この上清 5μl に PCR プライマー M4 (5'-GTTTT CCCAG TCACG ACGTTG-3') 5pmol, RV (5'-CAGGA AACAG CTATG ACCATG-3') 5pmol を加え最終量 25μl で 96°C, 10 分間置く。72°C に温度を下げた後、2× reaction buffer [40mM Tris HCl(pH8.9), 3mM MgCl₂, 50mM KCl, 200μg gelatin/ml, 200μM each of dNTP mixture] 25μl, 1.25units Taq DNA polymerase を加え、PCR 反応を行なった。PCR 反応は、96°C 30 秒間, 55°C 1 分間, 72°C 2 分間の反応を 35 回繰り返して行なった。反応後、PCR 反応液を millipore フィルター (VS 0.050μm) の上に乗せて、TE [10mM Tris HCl(pH8.0), 1mM EDTA] に対して透析を 1 時間行なった (drop-dialysis)。

透析したサンプルを水で 2 倍量に薄め、サイクルシーケンス反応に用いた。A(アデニン)については、薄めたサンプル液 2μl に reaction mixture [160mM Tris HCl(pH8.9), 10mM MgCl₂, 50mM (NH₄)₂SO₄, 50μM dATP, 12.5μM dCTP, 75μM 7-deaza-dGTP, 50μM dTTP, 0.8unit Taq DNA Polymerase, 0.4pmol FAM T7 Primer (5'-dye-TAATA CGACT CACTA TAGGG-3')] 3μl を加え、95°C 30 秒間, 60°C 1 分間, 70°C 2 分間の反応を 15 回繰り返し、続けて 95°C 30 秒間, 70°C 1 分間の反応を 15 回繰り返した。C(シトシン)、G(グアニン)、T(チミン)についても同様

にJOE, TMRA, ROX プライマーを用いて、反応を行なった。なお、A, C, G, T の反応液量の比は5:5:10:10である。反応後、4つの反応液を集め(全部で30 μ l)、エタノール沈殿を行ない、4 μ l の loading buffer [formamide/50mM EDTA (5/1 by Vol/Vol)] に溶かし、373A DNA オートシーケンサー(Ver1.0.1., Applied Biosystems) を用いてシーケンスを決定した。シーケンス反応は5'側のプライマーを用いて行なった。

データ解析

得られた cDNA シーケンスデータはクローニングサイトのシーケンスと Poly A を除いた後、得られたシーケンス同士で互いに Similarity search を行ない、その後、DNA databank GenBank(r.72) に対して Similarity search を行なった。用いたプログラムは Blastn である⁹⁾。90% 以上の Similarity が存在したとき同定と見なした。GenBank に対して Similarity の見つからなかったクローンについて Protein database PIR(r.34) に対して Homology search を行なった。用いたプログラムは Blastx である⁹⁾。Blastx は核酸塩基配列を自動的に3つのフレームでアミノ酸に翻訳し、Homology を検出することができる。さらに、未知のクローンについて ORF(Open Reading Frame) 検索を行ない、cDNA の5'側から最初に現われた ATG サイトからアミノ酸に翻訳し、そのアミノ酸配列の Hydrophobicity について調べた。用いたアプリケーションは DNASIS(Ver.2.0) である。また、motif Database Prosite に対しても検索を行なった。用いたアプリケーションは Mac Pattern(Ver.2.0) である。これらのデータ解析は Sparc Station II(SunMicroSystem) と Macintosh II を用いて行なった。

データの流れを (Fig. 8) に示す。

[結果]

両末端 cDNA のサイズ

まず、全長 cDNA ライブラリーの長さを調べ、その cDNA ライブラリーの質を確かめたあと、MboI で切り、セルフライゲーションを行い両末端 cDNA ライブラリーを作製し、その cDNA インサートが PCR 法で増幅可能な長さかどうかを調べた。

HepG2 細胞から調製した cytoplasmic mRNA から全長 cDNA を合成しプラスミドベクター(pSPORT1)にライゲーションした後、MboI (GATC) を用いて切る前に、この cDNA をコンピテントセル DH5 にトランスフォーメーションし 1.5 ml の LB 培地で培養した後、rapid preparation でプラスミド DNA を調製した。もし、元の全長 cDNA ライブラリーが短い長さのクローンばかりであったとすると、この全長 cDNA ライブラリーの中で本当に全長を持つクローンはほとんどないことになる。調製したプラスミドを NotI, EcoRI で切り cDNA インサートの長さを調べた。全長 cDNA ライブラリーの平均の長さは約 2kbp であった。mRNA の平均の長さが約 2 kbase といわれていることから、この全長 cDNA ライブラリーの質は良いものであるといえる。しかし、このままでは PCR 法によって cDNA インサートを増幅できないので、大量かつ迅速な塩基配列決定法は適応できない。

次に、MboI (GATC) を用いてライゲーションしたすぐ後の全長 cDNA を切り、DNA 濃度を薄めてセルフライゲーションして両末端 cDNA ライブラリーを作製した。このときベクターのシーケンスはメチレーションによって保護されており MboI で切れないで、合成した cDNA インサートの部分のみが切れる。その後、DNA 濃度の薄い条件でセルフライゲーションする。このようにして両末端 cDNA ライブラリーが完成する(Fig.1 (a), (b))。両末端 cDNA をコンピテントセル *E.coli* DH5 にトランスフォーメーションした後 125 μ l の LB 培地で 24 時間培養した後、直接 PCR 法によってプライマー M4/RV を使って cDNA インサートを増幅した。PCR 産物の長さを測ったところ、0.3 kbp から 1.6 kbp まで様々であるが、平均約 800 bp であった(Fig. 2)。cDNA をこのような長さに短くすることによって PCR ダイレクトシーケンス法を用いて大量かつ迅速な塩基配列決定ができる。なお、PCR で使用したプライマー (M4/RV) とクローニングサイトのシーケンスの長さの和が約 200 bp あるので、cDNA の平均の長さは実質 600 bp となる。よって、この両末端 cDNA ライブラリーは 3' 末端の Poly A から上流へ平均約 300 bp に MboI サイトがあり、

それからさらに平均約 300 bp のところに 5' 末端が来る構造を持つと予想される。

このように大腸菌からの直接 PCR 法によって cDNA インサートを増幅できる様な適当な長さを持ち、cDNA の 5' 側と 3' 側の両方を持つ、真ん中の部分を取り除いた両末端 cDNA ライブラリーが完成した。

シーケンスエラー

未知のクローンに対してアミノ酸へ翻訳し機能的な予測をするためには、なるべくシーケンスエラーが少ないほうがよい。どのくらいの長さまで、不確定な塩基がほとんどなくシーケンスを決定できるかを調べた。

この両末端 cDNA ライブラリーの中からランダムに 275 クローンを選び、シーケンスを決定した。シーケンス反応は 5' 側の T7 dye プライマーを用い、PCR の反応液を透析したものをそのままテンプレートとして、サーマルサイクルシーケンス法を使って行った。373A オートシーケンサー(ABI)を用いてシーケンスを決定した。シーケンスのエラーについて解析した結果、cDNA の 5' 末端より 100 base までが 1/100(1%)、300 base までが 8/300(2.7%)であった。350 base よりも長い領域では不確定な塩基が急に多くなるのでアミノ酸への翻訳とその解析には適さない。この結果、この cDNA ライブラリーの 5' 末端から約 300 base 以下のシーケンスは、アミノ酸へ翻訳し、タンパク質データベースと Homology search を行うことが有効であるといえる。

Expression Profile

シーケンスを決定した 275 クローンについてどのような遺伝子がどのくらいの頻度で発現しているかを調べるためにそれぞれのシーケンスをお互いに Similarity search を行いその後、GenBank(r.72) に対して Similarity search を行った。用いたプログラムは Blastn である。その結果、91 クローン(33%)が既知の遺伝子であった(Table 1)。また、275 クローンは 251 種類に分類できた。このライブラリーと同じ HepG2 細胞より作製した 3'-directed cDNA ライブラリー(HepG2- 3'-directed cDNA ライブラリー)を GenBank(r.71) に対して Similarity search を行った時は既知のものが 22% であったが、これは DNA データベースの update によって登録されている遺伝子の数が飛躍的に増えたことによって既知の遺伝子と同定される確率が高くなったと考えられる。また、Expression profile について調べてみると、発現頻度の多いクローンほど既知の遺伝子である確率が高い。一番多く発現している遺伝子は elongation factor で、2 番目が

serum albumin であった。

発現している遺伝子の種類について解析すると、expression profile と発現量の多い遺伝子は HepG2 細胞における mRNA レベルでの特徴をよく表している (Fig.3)。例えば、abundant なクローンの中に elongation factor, albumin, ribosomal protein など HepG2 細胞でよく発現している遺伝子がある。また HepG2-3'-directed cDNA ライブラリーで同定された遺伝子と比較して、keratin18, calmodulin, alpha-fetoprotein, acidic ribosomal phosphoprotein P2, alpha-1-antitrypsin, 90kD heat shock protein など共通した遺伝子の発現が見られる²⁾(Table 1)。また、mRNA の発現量が多い遺伝子の種類がほぼ一致しており、クローンの数に対する頻度や発現している遺伝子の種類も似通っている。これらの結果より、HepG2-3'-directed cDNA ライブラリーとこの両末端 cDNA ライブラリーがよく似た遺伝子発現パターンを表しているといえる。同じ HepG2 細胞から作製した cDNA ライブラリーであるので当然であるかもしれないが、この両末端 cDNA ライブラリーは、元の全長 cDNA ライブラリーの作製の際に cDNA をカラムに通してサイズフラクションを行い、長い cDNA フラグメントのみを集めるというバイアスがかかると思われる方法を使っている。それにもかかわらず、2つのライブラリーは同じような発現パターンを示しているので、3'-directed cDNA ライブラリー同様、この両末端 cDNA ライブラリーもほとんどバイアスのないものであるといえる。従って、この cDNA クローンの発現頻度は、直接、その細胞での mRNA の発現頻度を表していると考えられる。

両末端 cDNA ライブラリーの評価

この両末端 cDNA ライブラリーの中でどれくらいの割合のクローンが正確に 5' 末端と 3' 末端の両末端をもっているかどうかを調べるために、つまりこの両末端 cDNA ライブラリーの質を評価するために、GenBank(r.72) に対して同定された 91 クローンについて解析を行った。もし、この cDNA が期待通りにできているとすると、cDNA の 5' 末端から *Mbo*I サイトまでのシーケンスがデータベース (GenBank) に登録されている遺伝子の mRNA シーケンスの 5' 末端から最初に現われる *Mbo*I サイトまでと一致し、かつ、cDNA の *Mbo*I のサイトから 3' 末端までがその mRNA シーケンスの 3' 末端に一番近い *Mbo*I サイトから 3' 末端までと一致する構造を持つ (correct structure)。cDNA の 5' 側が mRNA の 5' 側であるという判断は、開始コドン (ATG) を含んでいるか、あるいはそれより上流のシーケンスを含んでいれば 5' 末端をもったクローンであるとした。

91 クローン中 34 クローン(37%) が mRNA の 5' 末端と 3' 末端の両末端を含む期待通りの構造を持っていた(Fig.4, A-1,A-2)。残りの 57 クローンのうち 44 クローン(48%) は mRNA の 3' 末端を含むが、5' 末端を含まない構造の cDNA であった(Fig.4, B-1,B-2)。つまり、cDNA の 5' 末端から *Mbo*I サイトまでのシーケンスが完全な遺伝子の mRNA シーケンスの途中からそのすぐ下流にある *Mbo*I サイトまでと一致し、かつ、cDNA の *Mbo*I サイトから 3' 末端までがその mRNA シーケンスの 3' 末端に一番近い *Mbo*I サイトから 3' 末端までと一致する構造を持つものであった。このような不完全な構造を持つ cDNA クローンができた理由として、全長 cDNA ライブラリーを作製するときに 1 st.strand 合成が途中で止ったか元の mRNA が分解していたために生じたものと思われる。また、cDNA の *Mbo*I サイトより 3' 側のシーケンスが mRNA の 3' 側と一致するが 5' 側が未知のシーケンスであるクローンや、その逆で、cDNA の *Mbo*I サイトより 5' 側のシーケンスが mRNA の 5' 側か真ん中と一致するが 3' 側が未知のシーケンスであるクローンが 13 クローン見つかった(Fig.4, C-1,C-1,C-2)。これらは chimeric なクローン、あるいは全長 mRNA シーケンスがデータベースに登録されていない、あるいは alternative splicing された mRNA の cDNA から生じたものと思われる。C-1 についてはまだ GenBank 登録されていない、さらに上流の 5' 末端をもったクローンである可能性も考えられる。

次に、この両末端 cDNA ライブラリーの 5' 側がどれくらいの割合で coding シーケンスを含んでいるのかを調べた。coding 領域を含むということは未知のクローンをアミノ酸に翻訳することによって機能的な予測を可能にする。GenBank に対して同定された 91 クローンの中で coding シーケンスを含むものが 74 クローン(81%) あり、そのうち 35 クローン(38%) が翻訳のための開始コドン(ATG)を含み、タンパク質の N 末端の情報を持ったクローンであった。このことより 3'-noncoding region と比べて 5'-noncoding region の長さが短いといえる。

これらの結果、特にサイズの大きなクローンを除き、作製した両末端 cDNA ライブラリーの中では全体の約 4 割のクローンが mRNA の 5' 末端と 3' 末端の両末端を持った期待通りの cDNA であると予想される。この 4 割という値はランダムに cDNA のシーケンスを決定し、解析するうえでかなり有用であると考えられる。

3'-directed cDNA ライブラリーとの Similarity

3'-directed cDNA ライブラリーはすでに様々な細胞、組織から作製され、大量にシーケンスが決定されており、それらの解析により、その細胞における遺伝子発現レベルでのより確かな発現頻度情報を持っている。この情報と両末端 cDNA ライブラリーの情報をリンクさせることによって、未知の遺伝子の様々な細胞での発現頻度情報が得られる。逆に考えれば 3'-directed cDNA に coding 含んだ 5' 側のシーケンス情報を付け加えることができる。

GenBank に対して Similarity を調べた結果同定されなかった未知の 184 クローンについて、3'-directed cDNA ライブラリーのシーケンスデータ(3664 クローン)に対して Similarity search を行った。用いた 3'-directed cDNA ライブラリーは HepG2 細胞から作製した HepG2-3'-directed cDNA ライブラリー 982 クローン、HL60 細胞(ヒト前骨髄性白血病由来の細胞株)から作製した HL60-3'-directed cDNA ライブラリー 1268 クローン、HL60 細胞に DMSO(dimethyl sulfoxide) 処理を行い分化誘導させた細胞から作製した HL60d-3'-directed cDNA ライブラリー 1417 クローンの合計 3664 クローンである。その結果 20 クローンの cDNA の *Mbo*I サイトから 3' 側のシーケンスがもうすでに 3'-directed cDNAs として登録されていた(Table 2)。HepG2-3'-directed cDNA ライブラリーに同定されたものが 10 クローン、HL60-3'-directed cDNA ライブラリーに同定されたものが 5 クローン、HL60d-3'-directed cDNA ライブラリーに同定されたものが 5 クローン見つかった。それぞれのクローンが各細胞でどれくらいの頻度で発現しているかがわかる。それらの cDNA をアミノ酸に翻訳して Protein Database PIR(r.34) に対して Homology search を行ったところ、Rat Ribosomal protein L9 とアミノ酸レベルで 100% の Homology を持っているものが見つかった。これは新しい Human Ribosomal protein であることが予想される。このクローンは 3'-directed cDNAs の解析により HepG2 細胞で 6/982(0.6%) の発現頻度を持つ abundant なクローンでしかもその性質が同定できないものであった。このほか、*C.elegans* Heat shock protein 70F や Mouse Transplantation antigen P198, Mouse MER5 protein などのタンパク質と Homology を持つものが見つかった。このように、3'-directed cDNA ライブラリーとの比較により新たな情報を与えることができた。

Protein Database PIR との Homology

GenBank に対して同定されなかった未知の 184 クローンについて、その機能を予測するために、3 種類のフレームで核酸配列からアミノ酸配列への変換を

行い、Protein database PIR(r.34) に対して Homology search を行った。用いたプログラムは Blastx である。その結果、アミノ酸レベルで 50% 以上の Homology を持つクローンが 19 クローン見つかった (table 3)。Rat Ribosomal protein L9, L35 などと強い Homology (90% 以上) を持つクローンが見つかった。それらは新しい Human Ribosomal protein であると考えられる。80% 代の Homology として Mouse Gene I3 protein, Bovine Phosphocarrier protein precursor, C.elegans Heat shock protein 70F, Human Tat-binding protein が見つかった。また、Rat Hemoprotein H-450 precursor と 51% の Homology を持つクローンについては、開始コドンと思われる ATG サイトがそのタンパク質の頭のアミノ酸であるメチオニンと一致していた。その他様々な生物のタンパク質と Homology を持つクローンが見つかった。これらは同様な機能を持つ未知の Human の遺伝子であると考えられる。

この両末端 cDNA の 5' 側、3' 側のどちらで Homology が見つかったを調べてみると、cDNA の 3' 側のみ、あるいは 5' 側から 3' 側へ両方にまたがって Homology を持つクローンも一部見られたが、ほとんどのクローンは cDNA の 5' 側のみで Homology を持っており、それらのクローンは 3' 側のみの解析から機能的な情報が得られなかったクローンである。このことから、ほとんどの両末端 cDNA クローンの 5' 側は coding 領域を含んでいると推測できる。これらクローン以外にも、40% から 50% の Homology を持つものが 17 クローン見つかった。

このように、GenBank に対して同定されなかった未知の cDNA ライブラリーシーケンスをアミノ酸に翻訳し Protein Database PIR と Homology search を行うことにより、いくつかのクローンについて機能的な予測をすることができた。

ORF 検索

さらに、機能的な予測を行うため、未知のシーケンスを持つ 184 クローンについて ORF (Open Reading Frame) 検索を行った。ORF 検索を行うことによって、Protein Database PIR と Homology が全く見られなかったクローンの塩基配列がどのフレームでアミノ酸に翻訳されているのかを予測することができる。3 種類のフレームでアミノ酸へ翻訳し、開始コドンと終止コドンの位置を調べた。その結果、105 クローン (57%) について ATG を頭に含むフレームが見つかった。そのいくつかのパターンを (Fig. 5) に示す。一方、GenBank に対して同定された既知のクローンの解析によると、開始コドンを示す ATG を含むクローンは 91 クローン中 35 クローン (38%) である。この 2 つの割合 (57% と 38%)

はかなりの差であるが、実際は、未知のクローンの ORF 検索によって見つかった 105 クローンの ATG サイトのすべてが開始コドンとして働いているわけではないと考えられる。以上のように ORF 検索で見つかったフレームについてアミノ酸に翻訳しその配列に対する Hydropathy 解析とモチーフ検索を行った。

Hydropathy 解析

Hydropathic index を調べることによってそのタンパク質の性質がある程度わかる場合がある。例えば、分泌タンパクであればその頭のメチオニンから疎水性のアミノ酸が約 20 a.a. 以上にわたり並び、膜たんぱくであれば膜貫通ドメインとなる部分のアミノ酸が約 20 a.a. 以上にわたり疎水性のアミノ酸となる^{10),11)}。GenBank に対して同定されなかった未知の 184 クローンについて ORF 検索を行い、ATG サイトを頭に含むフレーム 105 クローンについてアミノ酸に翻訳し、KYTE and DOOLITTLE の Index Table (Table 4) を用いて Hydropathy 解析を行った。この Index Table を用いてアミノ酸配列のスコアを計算すると、プラスのスコアが疎水性のアミノ酸配列、マイナスのスコアが親水性のアミノ酸配列を示す。スコアを計算する際 window サイズは 10 a.a. で行った。その結果、頭から約 20 アミノ酸までの領域で疎水性の強いクローン(Fig.6 (a))、頭から 10 から 30 アミノ酸の領域で疎水性の強いクローン(Fig.6 (b))、さらに、疎水性の領域と親水性の領域が交互に周期的に現われたり、あるいはアミノ酸配列の真ん中付近で疎水性のアミノ酸が 20 a.a. 以上にわたって存在するクローン(Fig.6 (c)) などいくつかのパターンが見られた。これらのクローンは分泌タンパク質あるいは膜タンパク質の特徴を持つものであると考えられる。

モチーフ検索

未知のシーケンスの解析に関してはモチーフ検索も有効であろう。現在までに様々なモチーフが知られており、特にアミノ酸配列モチーフはそのタンパク質の機能を知るうえで重要である。GenBank に対して同定されなかった未知の 184 クローンについて ORF 検索を行い、ATG サイトを頭に含むフレーム 105 クローンについてアミノ酸に翻訳し motif database Prosite (685 entries) に対して検索を行った。その結果、はっきりとしたモチーフが見つかったのは 3 クローンのみであった(Fig.7)。Zinc Finger モチーフ、Leucine Zipper モチーフ、RGD モチーフである。Zinc Finger モチーフ、Leucine Zipper モチーフは DNA 結合タンパクに見られるモチーフであり^{12),13)}、この cDNA は未知の Human の

DNA 結合タンパク質をコードしている遺伝子であると考えられる。また、RGD モチーフは fibronectin などに見られるモチーフで細胞間接着因子と深く関与しており¹⁴⁾、これらのクローンは未知の Human の細胞間接着タンパク質をコードしている遺伝子であると考えられる。

[考察]

両末端 cDNA ライブラリーの作製とその解析を行った。その結果、約 4 割のクローンが期待どおりに mRNA の 5' 末端と 3' 末端の両末端を持ち、cDNA の真ん中の部分の抜け落ちた構造(correct structure)を持つものであった。残りのクローンは mRNA の 5' 末端を持たない、mRNA の途中の部分を持った cDNA であった。その原因は、この両末端 cDNA ライブラリーが全長 cDNA ライブラリーを *Mbo*I で切って作製したものであるため、その元の全長 cDNA ライブラリーの中に全長でない cDNA が含まれているからだと考えられる。従って、この場合、元のいわゆる全長 cDNA ライブラリーの中で本当に全長を持つクローンは、全体の約 4 割であると考えられる。この割合は、現在における酵素反応系の効率や全長 cDNA ライブラリー作製の技術を考えれば、様々な情報を得るうえでかなり有用であるといえる。

全長でない cDNA すなわち、正確に 5' 末端を含んでいないクローンができる理由として、元の mRNA があまりにも長いとき、あるいは mRNA の途中に強い 2 次構造を持っていた場合、その全長 cDNA は作製されにくいと考えられる。そこで、GenBank に対して同定された既知の 91 クローンについて元の mRNA の長さとその 2 次構造について解析を行った。91 クローンの中で correct 構造を持つ cDNA クローン(Fig.4, A-1,A-2)の元の mRNA の長さをデータベース(GenBank)より調べたところ、一番短い acidic ribosomal phosphoprotein P2 の 460 base から一番長い NAD(P)H menadione oxdoreductase の 2447 base まで様々で平均すると 1406 base であった。また correct 構造以外のクローン(Fig.4, B-1,B-2, C-1, C-2, C-3)についても元の mRNA の長さを Database より調べたところ、一番短い ribosomal protein S17 の 477 base から一番長い pM5 の 4086 base まで様々で平均すると 2245 base であった。次に、この mRNA の 5' 末端まで伸びないで途中で止ったと考えられるクローン(Fig.4, B-1,B-2)についてその止った付近の mRNA の 2 次構造を調べてみた。すると、止ったサイト、すなわち cDNA の 5' 末端に対応する mRNA サイトのすぐ上流の 15 base から 30 base の長さの領域でパインドローム様の構造が見られたクローンが 44 クローン中 19 クローン(43%) 見つかった。この部分的な構造によって 1st.strand 合成のとき反応が途中で止ると考えられ、5' 末端を含まない cDNA クローンができる理由の一部と考えられる。このような解析結果より、ライブラリーの質(cDNA が正確に 5' 末端を含んでいるかどうか)は mRNA の長さだけでなく、その塩基配列による影響も大きいと考えられる。この他、5' 末端を含んでい

ない原因として、mRNA に結合したタンパク質によるブロック、mRNA の分解等が考えられる。

これに対して、質の良い全長 cDNA ライブラリーを作製するためには、まず新鮮な材料を使うこと、つまり質の良い mRNA を調製しそれを使うことである。また、5' 末端を含んだクローンの割合をより多くするために、1st strand 合成のときに高い温度で活性のある reverse transcriptase を使う、あるいは formamide, dimethyl sulfoxide(DMSO) を加えることにより mRNA の 2 次構造を解いたり、mRNA の調製のとき 5' 末端についている CAP 構造を認識してそのみをクローニングする¹⁵⁾ などの方法が考えられる。CAP 構造を認識する方法は、正確に 5' 末端を持つ cDNA がクローニングできるので理想的である。いかにして質の良い全長 cDNA ライブラリーを作製するかが今後の課題である。

逆に、この両末端 cDNA ライブラリーの塩基配列が PCR ダイレクトシーケンス法を用いて簡単に、しかも大量かつ迅速に決定することができることを利用すると、全長 cDNA ライブラリーを作製したときの質を計るモニターとして使うことができる。つまり、ある細胞・組織から全長 cDNA ライブラリーを作製したとき、その cDNA ライブラリーの中で本当の全長 cDNA を持つクローンがどのくらいの割合で存在するかを調べるために、その全長 cDNA ライブラリーを適当な制限酵素で切りこのような両末端 cDNA ライブラリーを作製し大量シーケンス、解析を行う。そして正確に mRNA の 5' 末端がクローニングされているかどうかを調べることによって元の全長 cDNA ライブラリーの質が推測できる。

また、3'-directed cDNAs とこの両末端 cDNAs のデータをリンクさせることは重要である。3'-directed cDNA データが今後さらに大量に増加するにつれて、両末端 cDNA との Similarity search により同定される確率が高くなり、同定されれば、そのクローンがどのような細胞・組織でどのくらいの頻度で発現しているかという情報が得られる。逆に言えば、その 3'-directed cDNA クローンは機能的情報を持った 5' 側シーケンスが得られる。しかし、この両末端 cDNA ライブラリーの平均の長さは 600 bp であり、今日の Single run シーケンスで決定可能な長さ約 400 base よりも長い。実際、MboI サイトを越えて正確にシーケンスが決定できているクローンは 275 クローン中 179 クローン (65%) であった。すべてのクローンを Single run シーケンスで塩基配列を決定するためには、元の全長 cDNA ライブラリーを MboI 以外の制限酵素で切り、両末端 cDNA ライブラリーの平均の長さを 400 bp 以下にすることが考えられるが、cDNA の長さがあまりにも短いと機能的な予測が難しくなることや、メ

チレーションによるセレクションを使える制限酵素が限られていることなどにより、その方法は期待できない。むしろ、シーケンスの技術的改良により、例えばキャピラリー電気泳動を応用した DNA シーケンス法などを用いて¹⁶⁾、Single run シーケンスでより長い DNA フラグメントの塩基配列が決定できることが望まれる。

一方、この両末端 cDNA ライブラリーの 5' 側のシーケンスデータの蓄積によって、将来、mRNA の 5' 末端に特有な配列が見つかる可能性が考えられる。もしそうなれば、この cDNA ライブラリーから得られたデータは極めて有用である。さらにこの cDNA シーケンスのデータ解析で特に有用であるのは、ORF 検索を行い cDNA の 5' 側の最初の ATG サイトからアミノ酸に翻訳しその配列について hydropathy 解析を行うことである。この cDNA の 5' 側シーケンスの最初の ATG が開始コドンかどうかについては、Kozak M. が行った 5'-noncoding シーケンスの解析により、開始コドン ATG 付近にあるコンセンサスシーケンス(GCC[A/G]CCATGG) が知られており¹⁷⁾、これと比較することによってある程度予想することができる。ORF 検索によって見つかった 105 クローンの ATG サイト付近のシーケンスと Kozak のコンセンサスシーケンスを比較したところ、66 クローンがその ATG サイトの前後に似通った配列を持っていた。この割合は GenBank に対して同定されなかった未知のクローン(184 クローン)の 36% を占める。この結果からも、この両末端 cDNA ライブラリーの中で正確に 5' 末端を持つクローンの割合が約 4 割であるといえる。

このように、cDNA の 5' 側の解析は未知のシーケンスに対して機能的な予測をすることができる。さらにこの領域はシグナルペプチドなどのタンパク質の N 末端情報を含んでおり、ある細胞でのタンパク質 N 末端カタログを作ることができる。タンパク質のアミノ酸配列のシーケンスは N 末端から行われており、これらの情報と比較することにより、より有用なデータとなり得る。

一般に、5'-noncoding region の長さは 3'-noncoding region と比べて短く、その長さが 100 base 以下の mRNA が全体の約 70%、200 base 以下が 90% を占める¹⁷⁾。従って、mRNA の 5' 末端から約 300 base の領域のシーケンスはタンパク質の N 末端配列の情報を持っていると考えられる。仮に cDNA がある遺伝子の完全な mRNA の 5' 末端を持っていなくても、その遺伝子の coding 領域を持っていたならば、アミノ酸に翻訳することができる。アミノ酸配列が得られれば、タンパク質データベースとの Homology search やモチーフ検索によって未知のクローンに機能的な情報を与えることができる。そういう意味で、この両

末端 cDNA ライブラリーのほとんどのクローンはその 5' 側に coding シーケンスを含んでいると考えられ、未知のクローンの機能・性質を推測するための重要な情報を持っている。

さらに今後 PCR 法の技術改良により、より長いフラグメントが増幅可能となれば、この両末端 cDNA ライブラリーの解析によって興味あるクローンが見つかったとき、その 5' 側のシーケンスと 3' 側のシーケンスを用いてプライマーを合成し PCR 反応を行うことにより、簡単にその遺伝子の全長 cDNA が得られる。また、同一遺伝子の 5' と 3' のシーケンス情報が得られれば、未知のゲノムのシーケンスデータと比較することにより、その遺伝子の頭と尻尾の位置を決定することができる。

以上のように、この両末端 cDNA ライブラリーの解析は今後の遺伝子・ゲノムの解析を行う上で重要な情報を与えてくれるだろう。

[Fig., Table の説明]

Fig.1 (a) 両末端 cDNA ライブラリーの作製の原理と使用したプライマー

まず、全長 cDNA ライブラリーを作製しベクター pSPORT1 にライゲーションする。その後すぐに *Mbo*I(GATC) で cDNA の真ん中の部分を切り、DNA 濃度を薄めてセルフライゲーションを行う。*Mbo*I(GATC) はメチレーションサイトを切らないのでベクター部分は切れず、合成された cDNA の中だけが切れる。*Mbo*I は 4 base カッターなので理論的には平均 256 base に 1 箇所の割合で切れる。このようにして両末端 cDNA が完成する。cDNA は大腸菌に transformation され、ベクター内のプライマー M4, RV を用いて直接 PCR 反応を行い cDNA インサートを増幅する。T7 dye primer を使って PCR ダイレクトシーケンス法を用い PCR 法で増幅した cDNA の塩基配列を決定する。

(b) 両末端 cDNA を組み込んだベクター pSPORT1

cDNA インサートは pSPORT の *Not*I, *Sa*II サイトにクローニングされ、*Sa*II 側に cDNA の 5' 側が、*Not*I 側に cDNA の 3' 側が来る。

Fig.2 cDNA インサートの PCR 産物

cDNA インサートを大腸菌からの直接 PCR 法を用いて増幅させ、1.5% アガロースゲルでそのサイズを見たもの。PCR 産物の平均の長さは約 800 bp であるが、PCR プライマー(M4/RV) とベクタークローニングサイトのシーケンスの長さの和が約 200 bp あるので、cDNA の平均の長さは実質 600 bp となる。サイズマーカーは 1 kb ladder (BRL)。

Fig.3 expression profile and abundant gene list

275 クローンのシーケンスをお互いに Similarity search を行い、次に GenBank(r.72) に対して Similarity search を行った結果、251 種類に分類でき、91 クローンが同定された。黒く塗りつぶした部分は GenBank に対して同定されたクローンを表し、白抜きの部分は未知のクローンを表している。灰色の部分は未知のクローンの中で 3'-directed cDNAs に対して同定されたクローンである。abundant なクローンほど既知の遺伝子である確率がたかい。abundant な遺伝子のリストもあわせて示してある。一番多く見られた遺伝子は elongation factor でその頻度は 8/275(2.9%) であった。2 番目に多かったのは serum albumin でその頻度は 4/275(1.5%) であった。abundant な遺伝子は HepG2 細胞の特徴をよく表している。

Fig.4 両末端 cDNA ライブラリーの質

GenBank に対して同定された 91 クローンについて、その cDNA が元の mRNA のどの部分のシーケンスを持っているのかを示している。そのマッチングによって 5 種類に分類した。

(A-1) cDNA の 5' 末端から *Mbo*I サイトまでのシーケンスがデータベースに登録されている mRNA のシーケンスの 5' 末端から最初に現われる *Mbo*I サイトまでと一致し、かつ、cDNA の *Mbo*I サイトから 3' 末端までがその mRNA の 3' 末端に一番近い *Mbo*I サイトから 3' 末端までと一致する構造を持つ、しかも開始コドンである ATG サイトを含んでいるクローン。(5' 末端を含む)

(A-2) cDNA の 5' 末端から *Mbo*I サイトまでのシーケンスがデータベースに登録されている mRNA のシーケンスの 5' 末端から最初に現われる *Mbo*I サイトまでと一致し、かつ、cDNA の *Mbo*I サイトから 3' 末端までがその mRNA の 3' 末端に一番近い *Mbo*I サイトから 3' 末端までと一致する構造を持つ、しかし開始コドンである ATG サイトを含まないクローン。(5' 末端を含む)

(B-1) cDNA の 5' 末端から *Mbo*I サイトまでのシーケンスがデータベースに登録されている mRNA の途中からその下流にある最初の *Mbo*I サイトまでと一致し、かつ、cDNA の *Mbo*I サイトから 3' 末端までがその mRNA の 3' 末端に一番近い *Mbo*I サイトから 3' 末端までと一致する構造を持つクローン。(5' 末端を含まない)

(B-2) cDNA の 5' 末端から 3' 末端までが mRNA の 3' 側と一致する構造を持つクローン。cDNA のシーケンスの中に *Mbo*I サイトを含まないか、あるいは *Mbo*I サイトがあったとしてもその位置が 3' 末端に一番近い *Mbo*I サイトであるクローン。(5' 末端を含まない)

(C-1) cDNA の 5' 末端から *Mbo*I サイトまでのシーケンスがデータベースに登録されていない未知のシーケンスで、かつ、cDNA の *Mbo*I サイトから 3' 末端までがデータベースに登録されている mRNA の 3' 末端に一番近い *Mbo*I サイトから 3' 末端までと一致する構造を持つクローン。これらは alternative splicing された未知の mRNA から作られたものか、あるいは 5' 末端がまだ未知のクローンであると考えられる。

(C-2) cDNA の 5' 末端から *Mbo*I サイトまでのシーケンスがデータベースに登録されている mRNA の途中からその下流にある最初の *Mbo*I サイトまでと

一致し、かつ、cDNA の *MboI* サイトから 3' 末端までがデータベースに登録されていない未知のシーケンスである構造を持つクローン。(chimera)

(C-3) cDNA の 5' 末端から *MboI* サイトまでのシーケンスがデータベースに登録されている mRNA のシーケンスの 5' 末端から最初に現われる *MboI* サイトまでと一致し、かつ、cDNA の *MboI* サイトから 3' 末端までがデータベースに登録されていない未知のシーケンスである構造を持つクローン。(chimera であるが 5' 末端を含む)

(C-1), (C-2), (C-3) で '?' 印は同定されていない未知のシーケンスを表す。

Fig.5 ORF(Open Reading Frame) 検索

GenBank に対して同定されなかった未知のクローンの ORF 検索を行った。3 種類のフレームで開始コドンと終止コドンの位置を調べた結果、頭に ATG サイトがあるフレームを持つ cDNA が 105 クローン見つかった。そのうち 5 例を示す。▼ 印は ATG サイトを、■ 印は TAA, TAG, TGA サイトを示す。斜線部分は予想される ORF を示している。*MboI* サイトの位置もあわせて表示してある。

Fig.6 Hydropathy 解析

ORF 検索した結果、頭に ATG サイトを持つクローンについてアミノ酸に翻訳し KYTE and DOOLITTLE の Index Table を用いて Hydropathy 解析を行った。そのいくつかのパターンの結果を示す。横軸はアミノ酸配列を表し、縦軸はそのアミノ酸配列の Hydropathy index を表す。プラスが疎水性、マイナスが親水性を表す。window サイズは 10 a.a. で計算した。

(a) 頭のメチオニンから約 20 a.a. の領域で強い疎水性をしめすもの。分泌タンパク質の特徴を持っているものと思われる。

(b) 頭のメチオニンを 1 として 11 番目から 31 番目の間の約 20 a.a. の領域で強い疎水性を示すもの。これらも分泌たんぱく質の性質を持っているものと考えられる。

(c) 疎水性の部分と親水性の部分が周期的に現われるあるいはアミノ酸の真ん中の付近で疎水性の部分が 20 a.a. 以上にわたって存在するもの。膜タンパク質の特徴を持つものであると考えられる。

Fig.7 モチーフ検索

GenBank に対して同定されなかった未知のクローンについて motif Database

Prosite(r.7.0) に対して Mac Pattern を使いモチーフ検索を行った。その結果、3 クローンについてははっきりとしたモチーフが見つかった。Zink finger motif と Leucine zipper motif と RGD motif である。これらは、DNA 結合タンパク質や細胞間接着タンパク質が持っているモチーフである。

Fig.8 データ処理の流れ

シーケンスを決定した両末端 cDNAs 275 クローンのシーケンスデータの処理の流れを示す。これらの解析は Sparc Station II (Sun Mico System) と Macintosh II を用いて行った。

Table 1 GenBank に対して同定されたクローンリスト

91 クローンが GenBank に対して同定され、その遺伝子のリストを表している。Match region and *Mbo*I site のカラムは cDNA が mRNA のどの部分を持っているかを示している。なお、このカラムの中で '/' の部分は *Mbo*I サイトを示している。Size は元の mRNA の大きさを示している。一番右のカラムは、そのマッチングのパターンより Fig.4 で表しているように分類した場合のカテゴリを示している。

Table 2 3'-directed cDNAs との Similarity

GenBank に対して同定されなかった未知のクローンについて、3'-directed cDNAs に対して Similarity search を行った。その結果、20 クローンが同定された。全てのクローンで *Mbo*I サイトから 3' 側で同定されている。これらのクローンのそれぞれのライブラリー(HepG2 細胞、HL60 細胞、HL60 細胞を DMSO 処理して分化誘導した細胞から作製した cDNA ライブラリー)での発現頻度を同時に示している。5' 側の解析により、そのうちの 4 クローンがタンパク質データベースに対して Homology を持っていた。

Table 3 PIR との Homology

GenBank に対して同定されなかった未知のクローンについて、3 種類のフレームでアミノ酸に翻訳し Protein Database PIR に対して Homology search を行った結果、50% 以上の Homology を持つものが 19 クローン見つかった。Match site のカラムの '=' が Homology のあった領域を示している。また、cDNA の *Mbo*I サイトの位置により、あるタンパク質と Homology を持つ領域が cDNA の 5' 側か 3' 側のどちら側であるかがわかる。

Table 4 Hydrophathy index table (Kyte and Doolittle)

アミノ酸の疎水性の度合いを示す。プラスが疎水性、マイナスが親水性を表す。この表に従ってアミノ酸配列の Hydrophathy のスコアを計算する。

(The table content is extremely faint and illegible in the provided image. It appears to be a list of amino acids and their corresponding hydrophathy scores, as indicated by the caption and text.)

[参考文献]

- 1) Okubo, K., Hori, N., Matoba, R., Niiyama, T., and Matsubara, K. (1991). A novel system for large-scale sequencing of cDNA by PCR amplification. *DNA sequence* 2 137-144.
- 2) Adams, M.D., Kelley, J.M., Gocayne, J.D., Dubnick, M., Polymeropoulos, M.H., Xiao, H., Merril, C.R., Wu, A., Olde, B., Moreno, R.F., Kerkavage, A.R., McCombie, W.R., and Venter, J.C. (1991). Complementary DNA Sequencing: Expressed Sequence Tags and Human Genome Project. *Science* 252 1651-1656.
- 3) Okubo, K., Hori, N., Matoba, R., Niiyama, T., Fukushima, A., Kojima, Y., and Matsubara, K. (1992). Large scale cDNA sequencing for analysis of quantitative and qualitative aspects of gene expression. *nature genetics* 2 173-179.
- 4) Adams, M.D., Dubnick, M., Kerlavage, A.R., Moreno, R., Kelley, J.M., Utterback, T.R., Nagle, J.W., Fields, C. and Venter, J.C. (1992). Sequence identification of 2375 human brain genes. *Nature* 355 632-634.
- 5) Aden, D.P., Fogel, A. and Knowles, B.B. (1979). Controlled synthesis of HBsAg in a differentiated human liver carcinoma derived cell line. *Nature* 282 615-617.
- 6) Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989). Extraction and purification of RNA. *Molecular Cloning. 2nd.ed.*(New York: Cold Spring Harbor Laboratory) vol.1. pp 7.3-7.36.
- 7) Okayama, H. and Berg, P. (1982). High-efficiency cloning of full-length cDNA. *Mol.Cell Biol.* 2 161-170.
- 8) Hanahan, D. (1983). Studies on transformation of *Escherichia coli* with plasmid. *J. molec. biol.* 166 557-580.
- 9) Samuel, K and Altschul, S.F. (1990). Methods for assessing the statistical significance of molecular sequence features by using general scoring schemes. *Proc. Natl. Acad. Sdc.* 87 2264-2268.
- 10) Kyte, J. and Doolittle, R.F. (1982). A simple method for displaying the hydrophobic character of a protein. *J. Mol. Biol.* 157 105-132.
- 11) Engelman, D.M., Steitz, T.A. and Goldman, A. Identifying nonpolar transbilayer helices in amino acid sequence of membrane proteins. *Annu Rev. Biophys. Chem.* 15 321-353.
- 12) Klug, A. and Rhodes, D. (1987). 'Zinc fingers': a novel protein motif for nucleic acid recognition. *Trends in Biochem. Sci.* 12 464-469.

13) Landschilz, W.L., Johnson, P.F. and McKnight, S.L. (1988). The Leucine Zipper: A hypothetical structure common to a new class of DNA binding proteins. *Science* 240 1759-1764.

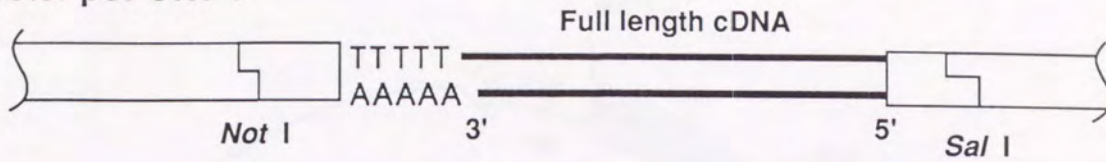
14) D'Souza, S.E., Ginsberg, M.H. and Plow, E.F. (1991). Arginyl- glycy- aspartic acid (RGD): a cell adhesion motif. *Trends in Biochem. Sci.* 16 246-250.

15) 菅野純夫、丸山和夫 (1993). 完全長 cDNA ライブラリーに向けて. *ゲノム解析研究*(共立出版). 蛋核酵 1993 年 2 月増刊.

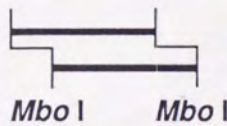
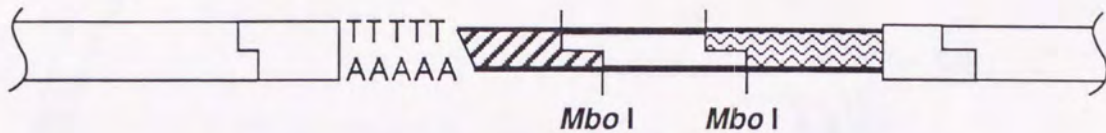
16) Mathies, R.A. and Huang, X.C. (1992). Capillary array elctrophoresis: an approach to high-speed, high-throughput DNA sequencing. *Nature* 359 167-169.

17) Kozak, M., (1987). An analysis of 5'-noncoding sequences form 699 vertebrate messenger RNAs. *Nucleic Acids Res.* 15 8125-8148.

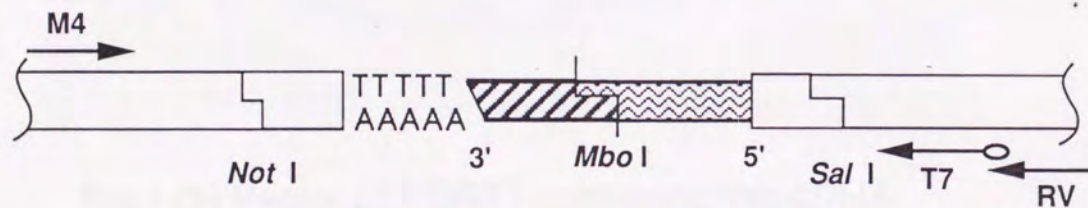
Vector pSPORT 1



Mbo I digestion



self ligation



PCR primer : M4 / RV

sequencing dye primer : T7

M4 : 5'-GTTTTCCAGTCACGACGTTG-3'

RV : 5'-CAGGAAACAGCTATGACCATG-3'

T7 : 5'-dye-TAATACGACTCACTATAGGG-3'

Fig.1 (a) Construction of a cDNA library containing both 5' and 3' ends

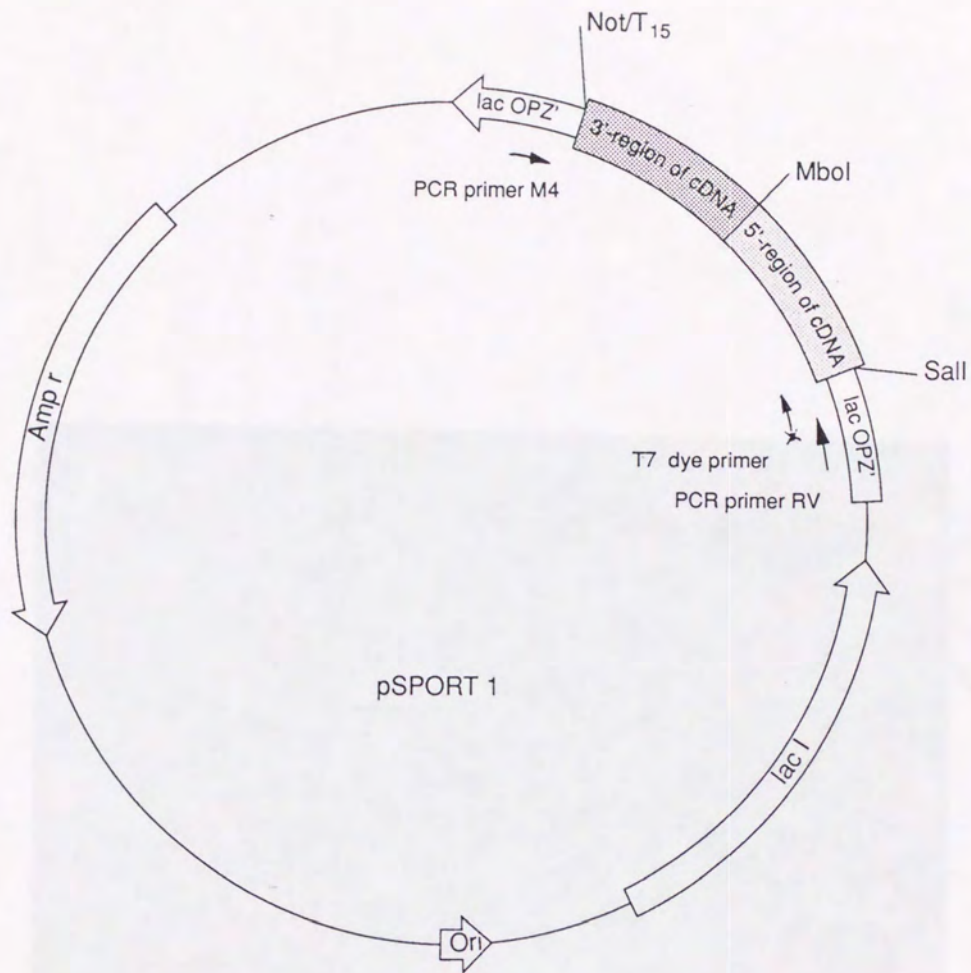


Fig.1 (b) Vector pSPORT1 containing this cDNA

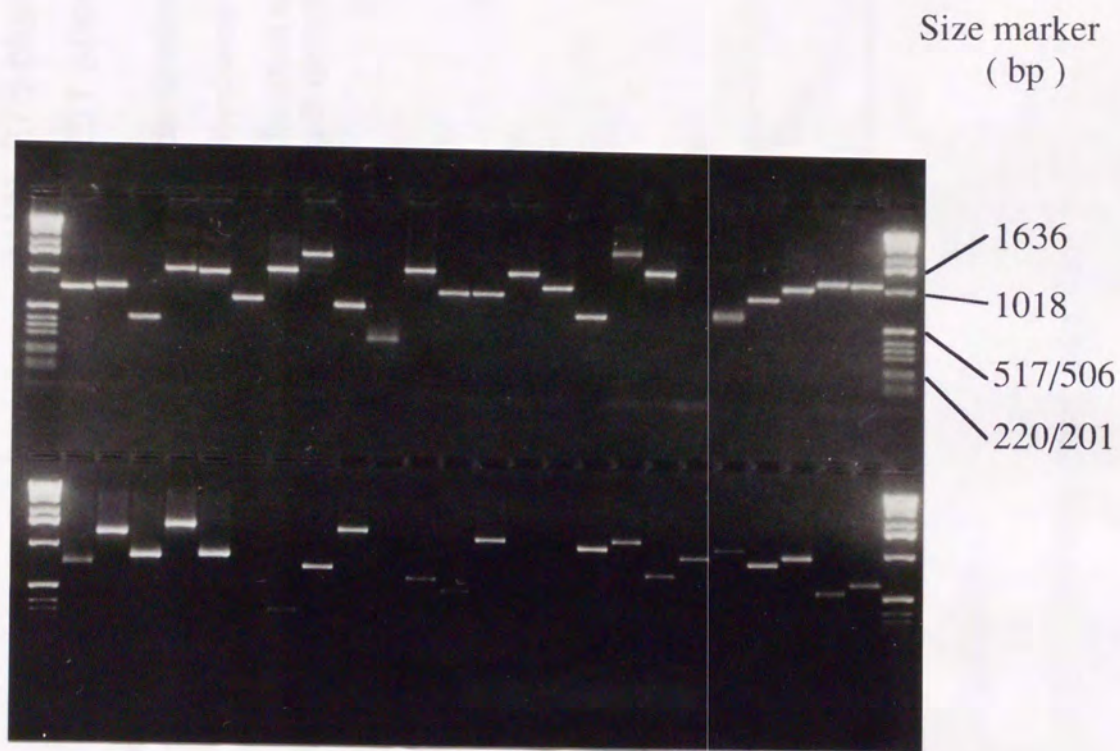


Fig.2 PCR products of this cDNAs

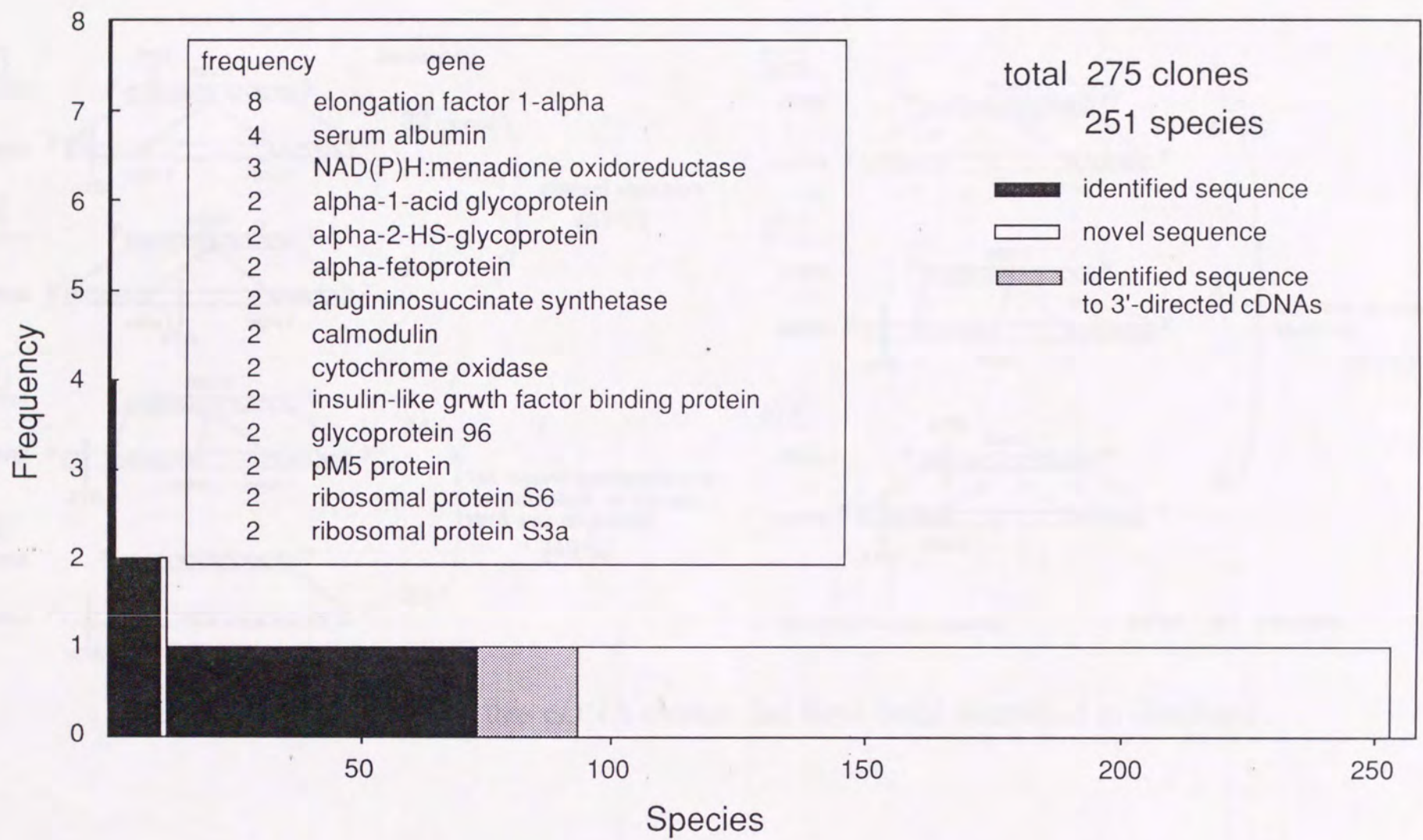


Fig.3 Expression profile and abundant gene list

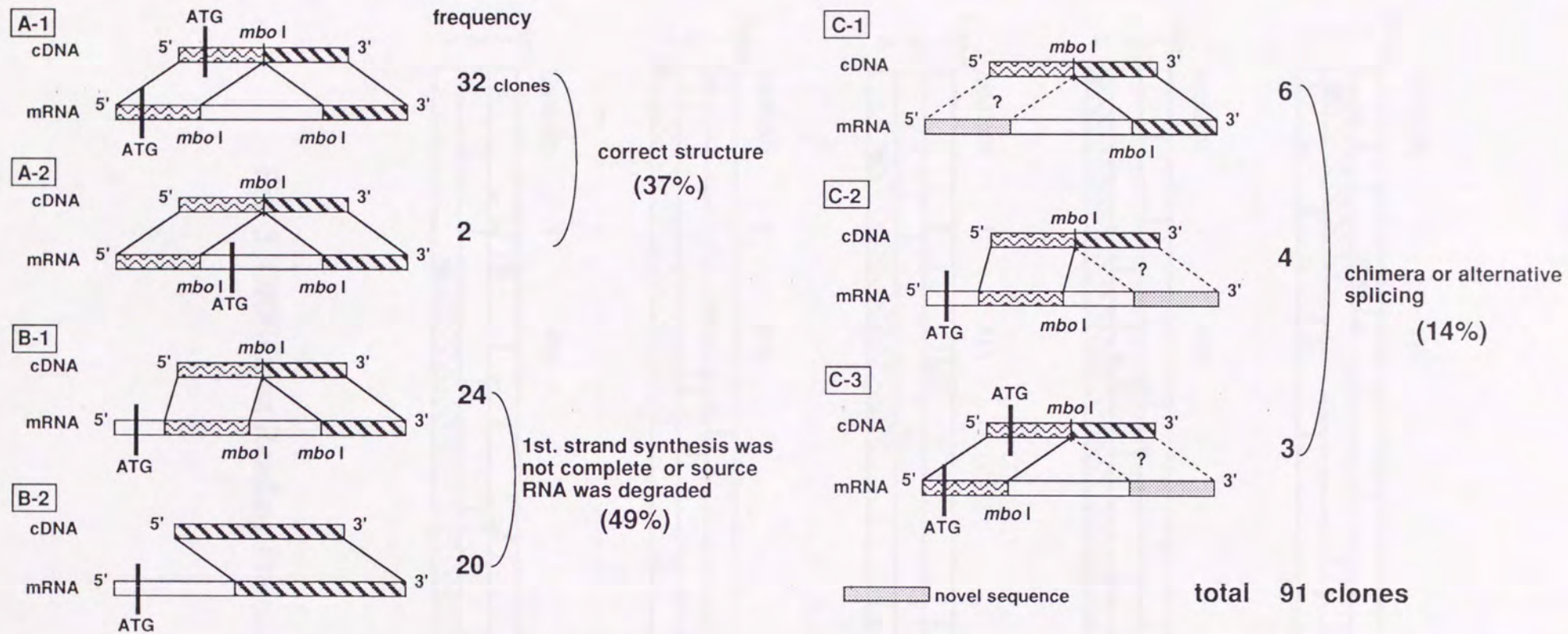


Fig.4 The structure of this cDNA clones that have been identified in GenBank

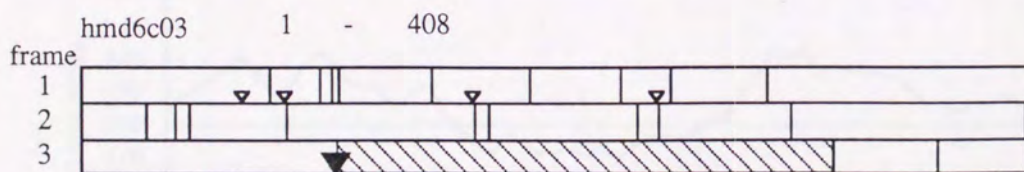
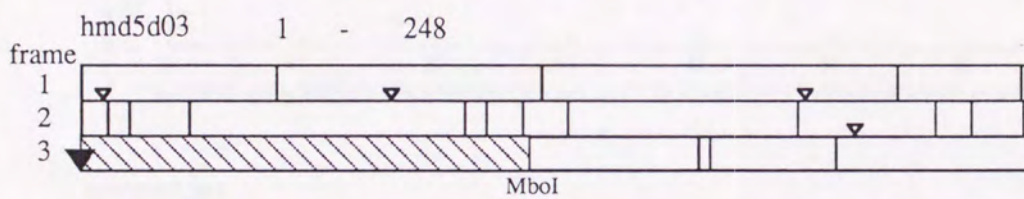
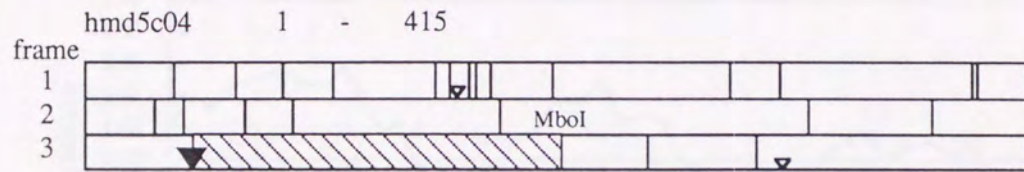
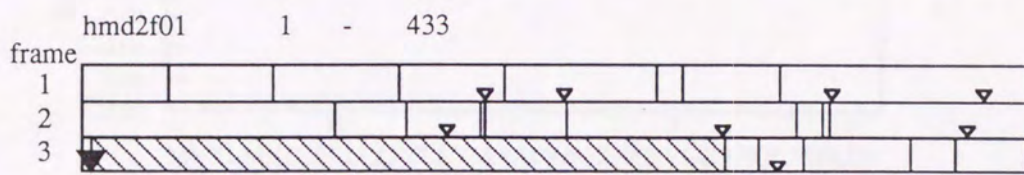
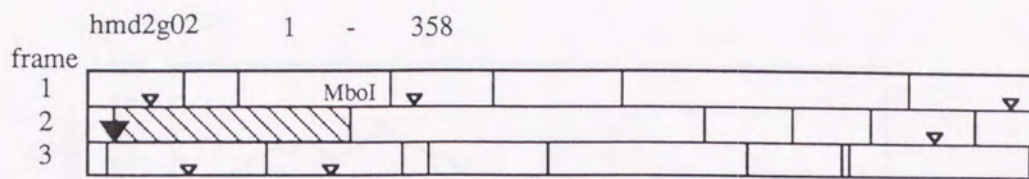


Fig. 5 ORF(Open Reading Frame) search

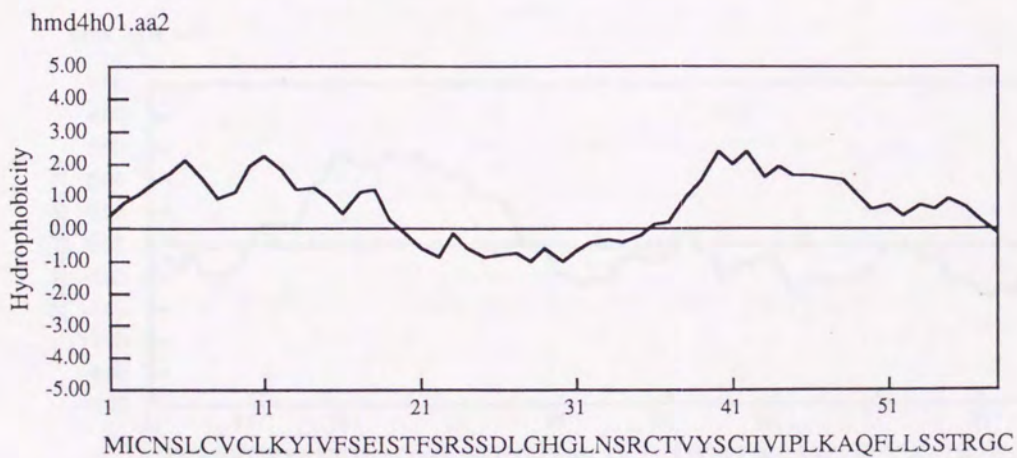
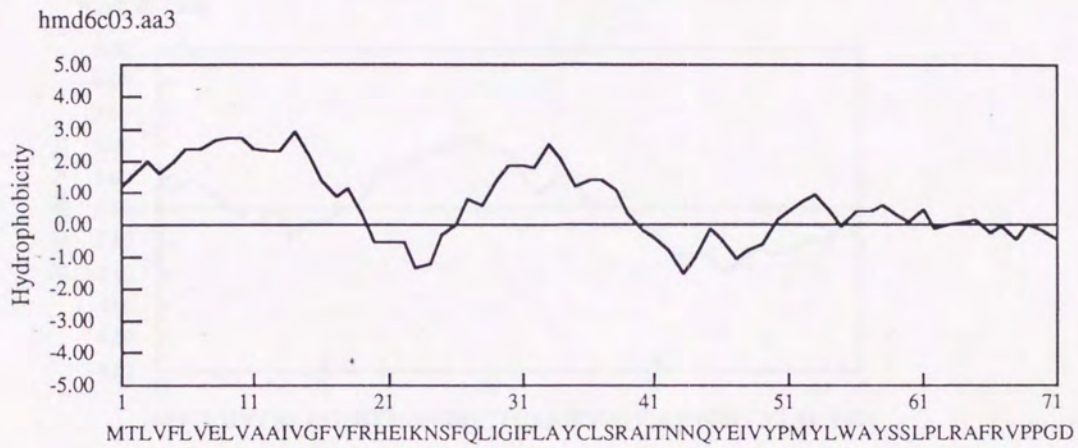
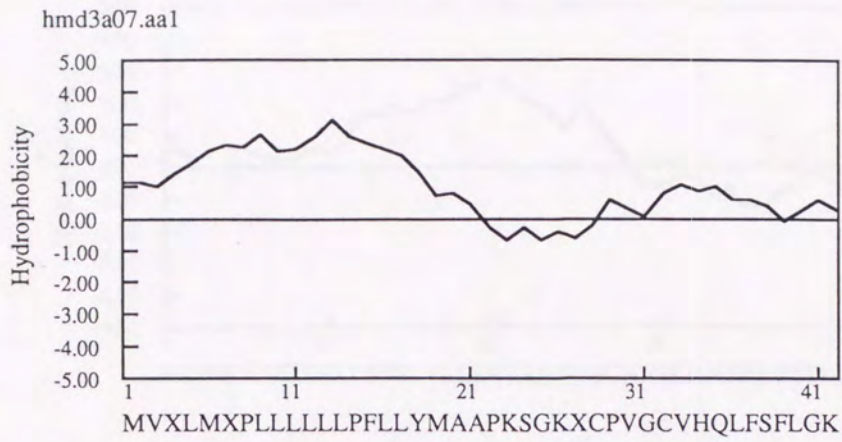


Fig.6 (a) Hydropathy analysis of a.a. sequences

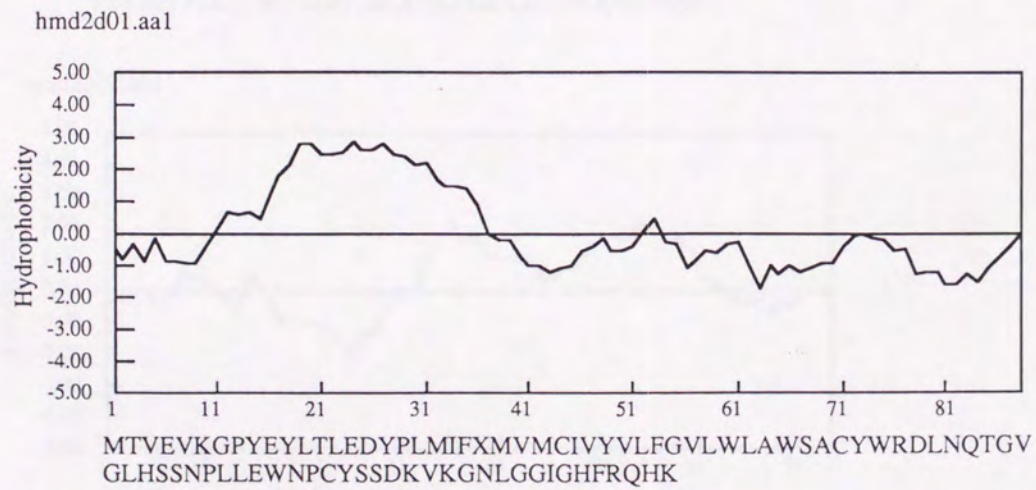
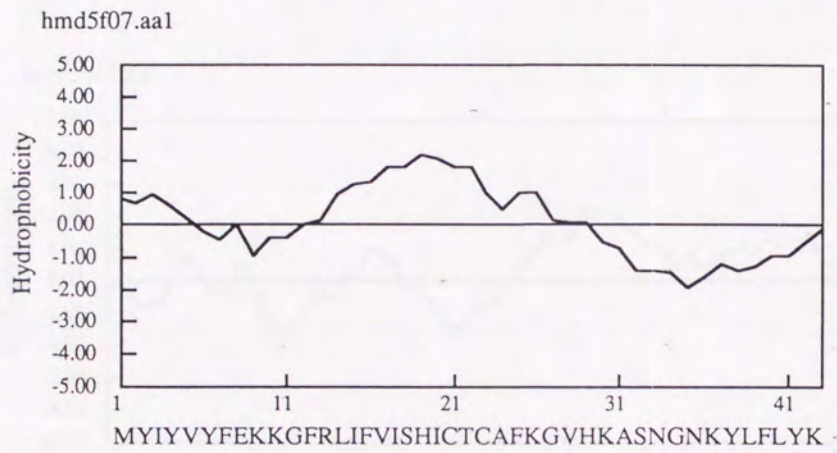
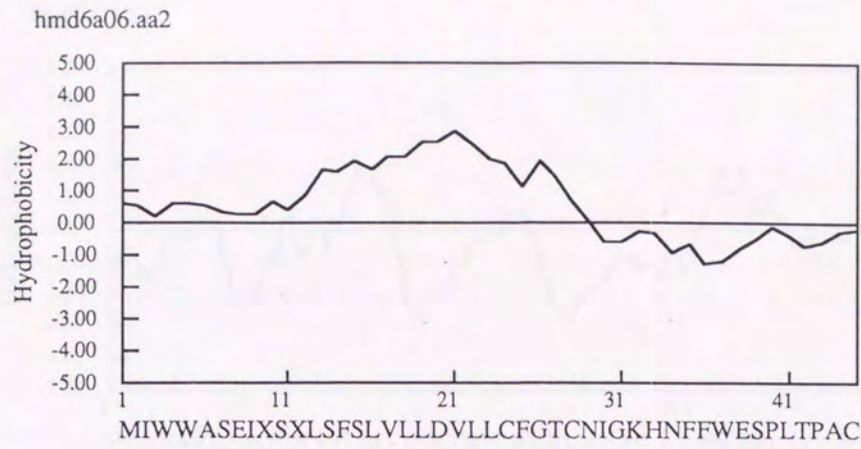


Fig.6 (b) Hydropathy analysis of a.a. sequences

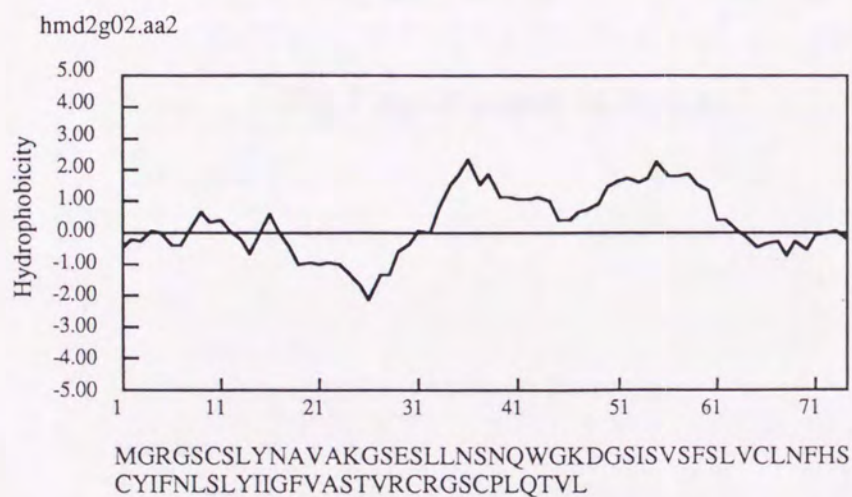
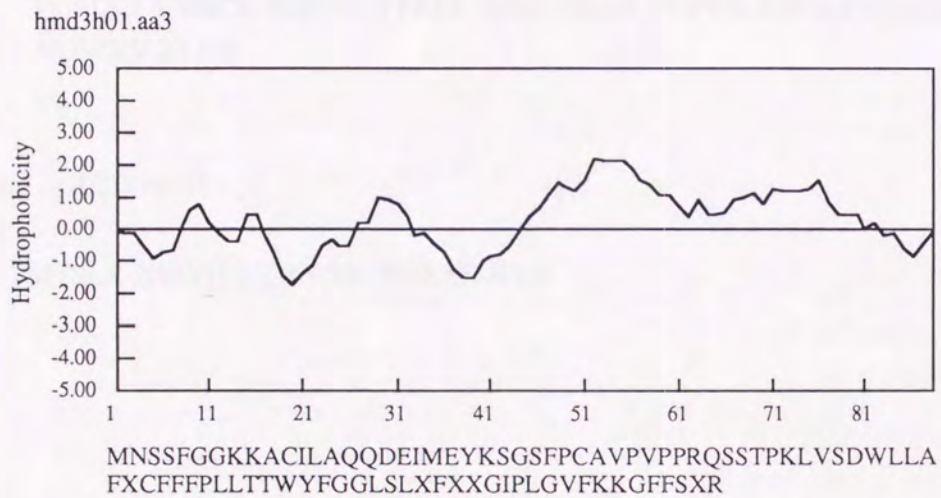
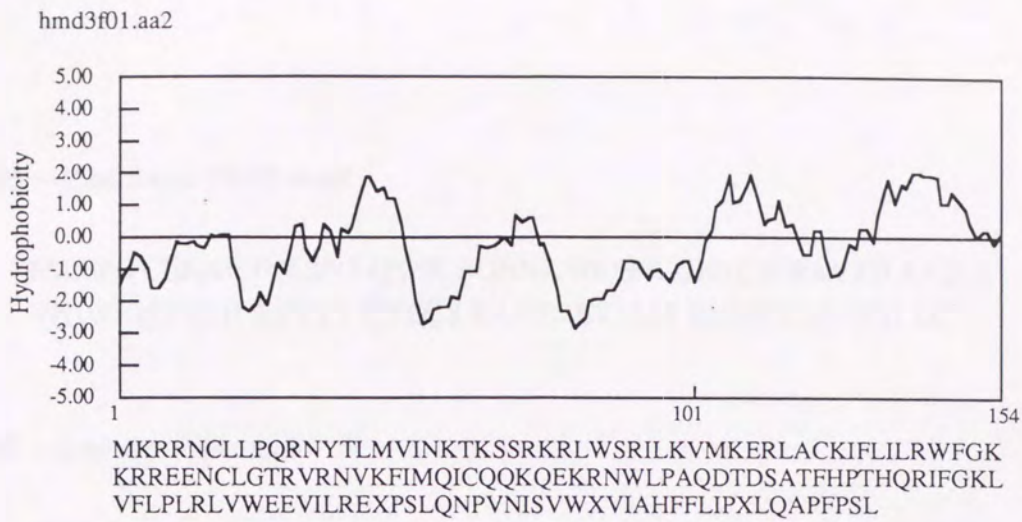


Fig.6 (c) Hydropathy analysis of a.a. sequences

hmd3g05 -- Zinc finger C2H2 motif

MAHPG CEKAF TQLSN LQSHR RQH NK DKPFK CHNCH RAYTD AASLE
VHLST HTVKH AKVYT CTECS RAYTS ETYLM KHMRK HNPPD LC

hmd2a05 -- Leucine zipper motif

MTQFL EQGEA TLSVA RRALA AQLPV LPRSE VTFLA PVTRP DKVVC
VGMXY VXHCK EQNVP VPKEP IIFSX FASSI VGPYD XVVLP PQSXX
VDWXV·XLGR

hmd5h03 -- RGD motif

MTSLR XWYRG DNPNK PMXXL IIYX

Fig.7 motif search in Prosite

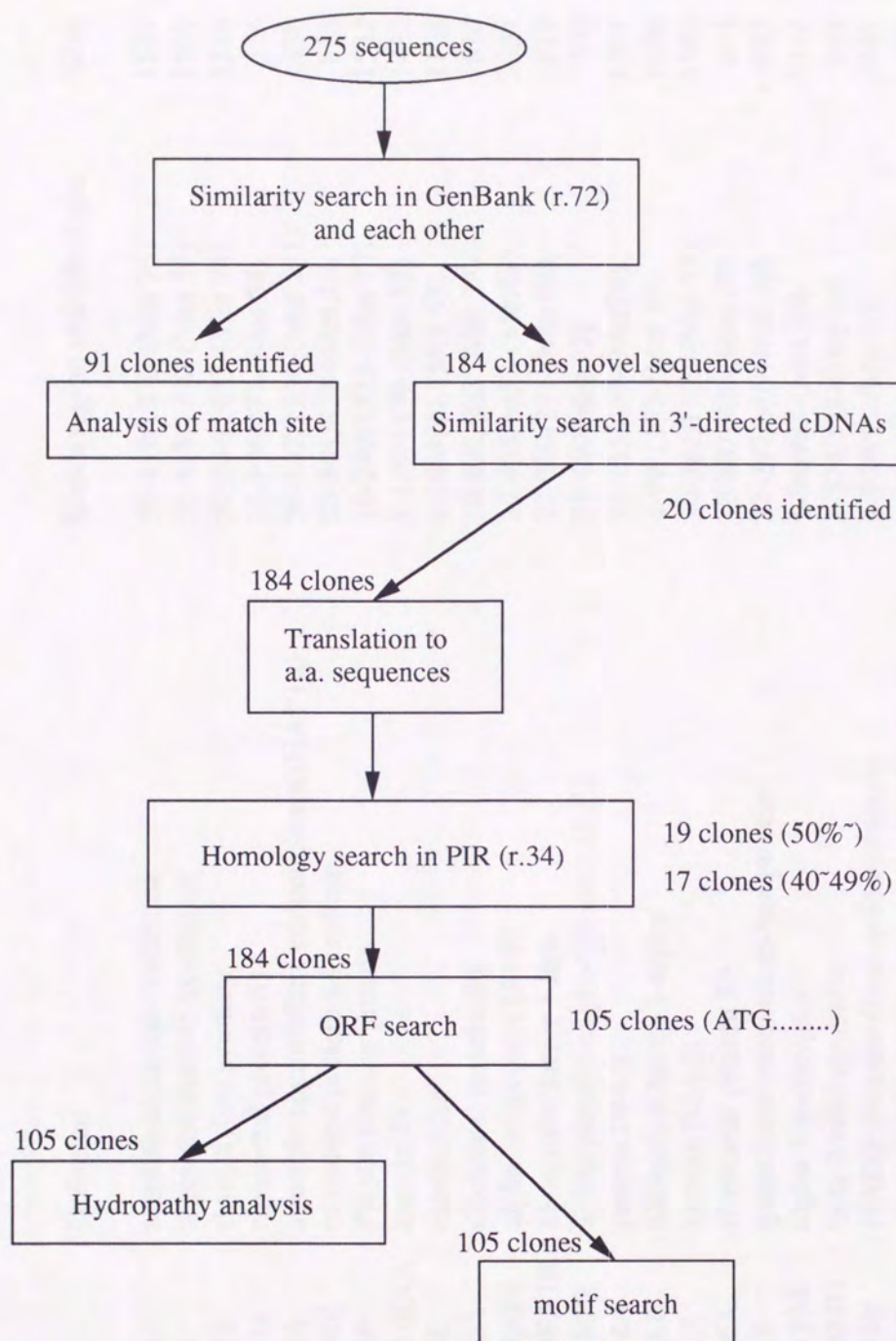


Fig.8 flowing of the Sequence Data

Table 1 Identified genes in GenBank (r.72)

cDNA clone	Locus	Defintion	Match region and MboI site	Size	Fig.4 Class	
1	hmd1a06	HUMASA	argininosuccinate synthetase	56-418/1370-(Met 76)	1547	A-1
2	hmd1b06	HUMASA	argininosuccinate synthetase	56-418/1370-(Met 56)	1547	A-1
3	hmd1c04	HUMTHYS	thymidylate synthase	50-204/1453-(Met 94)	1524	A-1
4	hmd1c08	HUMRPS14	ribosomal protein S14	33-196/477-(Met 58)	556	A-1
5	hmd1c09	HUMAPEA	apurinic/apyrimidinic endonuclease (HAP1h)	36-362/1134-(Met 211)	1438	A-1
6	hmd1d08	HUMRPS4X	ribosomal protein S4 isoform	25-341/852-(Met 36)	888	A-1
7	hmd1e02	HUMFKBP	FK506 binding protein	10-249/1375-(Met 79)	1532	A-1
8	hmd1f02	HUMKER18AA	keratin 18	5-189/1179-(Met 52)	1412	A-1
9	hmd2c09	HUMCAM	calmodulin	1-168/701-(Met 53)	1126	A-1
10	hmd2d02	HUMRPS6A	ribosomal protein S6	10-802/802-(Met 28)	814	A-1
11	hmd2e03	HUMALBAF4	alpha-fetoprotein (AFP)	22-455/1653-(Met48)	2032	A-1
12	hmd2f02	HUMELOFA1B	elongation factor 1-beta	73-369/791-(Met 96)	817	A-1
13	hmd2g01	HUMPPARP2	acidic ribosomal phosphoprotein P2	24-460(Met 75)	460	A-1
14	hmd2g11	HUMPAP4A	protein PP4-X	71-235/439-(Met74)	1301	A-1
15	hmd2h03	HUMEF1AC	elongation factor 1-alpha	1-99/1152-(Met 16)	1696	A-1
16	hmd2h09	HUMH2AX	histone H2A.X	50-382/1120-(Met 74)	1585	A-1
17	hmd3b09	HUMRPS6A	ribosomal protein S6	10-802/802-(Met 28)	814	A-1
18	hmd3c11	HUMGPLP	glutathione peroxidase-like protein	52-297/646-(Met 52)	985	A-1
19	hmd3e05	HUMA1ATM	alpha-1-antitrypsin	1-98/983-(Met 20)	1352	A-1
20	hmd3f02	HUMB2MMU	beta-2 microglobulin	1-374/538-(Met 16)	844	A-1
21	hmd3f10	HUMNMOR	NAD(P)H:menadione oxidoreductase	1-74/855-(Met 51)	2447	A-1
22	hmd4a01	HUMAHS	alpha-2-HS-glycoprotein alpha and beta chain	1-151/1265-(Met 49)	1538	A-1
23	hmd4g08	HUMAGP1A	alpha-1-acid glycoprotein	1-75/637-(Met 16)	737	A-1
24	hmd4h05	HUMEF1AC	elongation factor 1-alpha	1-99/1152-(Met 16)	1696	A-1

cDNA clone	Locus	Defintion	Match region and MboI site	Size	Fig.4 Class
25	hmd4h08	HUMRPL3A	ribosomal protein L3	1-180/880-(Met 7)	1256 A-1
26	hmd5b03	HUMAHSG	alpha-2-HS-glycoprotein alpha and beta chain	1-151/1265-(Met 49)	1538 A-1
27	hmd5e01	HUMALBAF4	alpha-fetoprotein (AFP)	23-452/1653-(Met 48)	2032 A-1
28	hmd6a12	HUMEFIAR	elongation factor 1-alpha subunit	19-137/1190-(Met 54)	1506 A-1
29	hmd6b05	HUMALB3	serum albumin	39-189/1924-(Met 40)	2055 A-1
30	hmd6b07	HUMRAP30	RAP30 subunit of transcription initiation factor RAP30/74	4-300/1357-(Met 120)	1408 A-1
31	hmd6b10	HUMGST	glutathione S-transferase	52-415/784-(Met 74)	909 A-1
32	hmd6f08	HUMGRANUL	H.sapiens granulin	1-300/1672-(Met 13)	2095 A-1
33	hmd4f09	HUMCAMA	calmodulin	20-72/1991-(Met 104)	2175 A-2
34	hmd5c04	HUMAGAGENE	aspartylglucosaminidase	1-123/1925-(Met 171)	2150 A-2
35	hmd1a10	HUMRPSA3A	ribosomal protein S3a 3' end	1-66/306-(Met <1)	836 B-1
36	hmd1b03	HUMNMOR	NAD(P)H:menadione oxidoreductase	1070-1563/2409-(Met 51)	2447 B-1
37	hmd1b05	HUMPMMP1	mitochondrial matrix protein P1	121-703/1505-(Met 46)	2227 B-1
38	hmd1c05	HUMTF	transferrin	1111-1227/2178-(Met 31)	2298 B-1
39	hmd1d12	HUMRIBIIR	ribophorin II	500-519/1968-(Met 289)	2509 B-1
40	hmd1f09	HUMNFKB	nuclear factor kappa-B	2766-2876/3584-(Met 398)	3625 B-1
41	hmd1g05	HUMPM5A	pM5 protein	1680-2043/4000-(Met 139)	4086 B-1
42	hmd2b04	HUMLSZ	lysozyme	3328-3405/5510-(Met 28)	5855 B-1
43	hmd2e06	HUMAPACT	alpha-actinin	1033-1050/1923-(Met <1)	2327 B-1
44	hmd2e07	HUMTRA1	glycoprotein 96 (tra1)	415-1310/2771-(Met 106)	2780 B-1
45	hmd2g03	HUMAPBB10X	apolipoprotein B-100	245-334/2744-(Met <1)	2811 B-1
46	hmd3c07	HUMSAP1	sphingolipid activator protein 1	666-798/1995-(Met 7)	2724 B-1
47	hmd3c10	HUMALB3	serum albumin	649-820/1924-(Met 40)	2055 B-1
48	hmd3e11	HUMMTSDHB	mitochondrial succinate-ubiqinone oxidoreductase iron sulfur subunit (sdh B)	1-218/865-(Met <1)	958 B-1
49	hmd3h02	HUMTRA1	glycoprotein 96 (tra1)	1352-1426/2771-(Met 106)	2780 B-1

cDNA clone	Locus	Defintion	Match region and MboI site	Size	Fig.4 Class
50	hmd3h04	HUMFGF1A	fibroblast growth factor (FGF) receptor-1	654-934/2688-(Met 10)	3343 B-1
51	hmd3h07	HUMADE2H1	H.sapiens ADE2H1 mRNA, homologies to SAICAR synthetase and AIR carboxylase of the purine pathway	714-856/1169-(Met 25)	1426 B-1
52	hmd3h08	HUMNMOR	NAD(P)H:menadione oxidoreductase	390-530/2409-(Met 51)	2447 B-1
53	hmd4b02	HUMALB3	serum albumin	364-820/1924-(Met 40)	2055 B-1
54	hmd4c10	HUMCYL	cyclin protein gene	353-569/1095-(Met 119)	1231 B-1
55	hmd5b04	HUMACTGAA	gamma-actin	791-1049/1049-(Met <1)	1210 B-1
56	hmd5c10	HUMPM5A	pM5 protein	3227-3266/4000-(Met 139)	4086 B-1
57	hmd5h11	HUMCTSB	cathepsin B proteinase	1121-1145/1878-(Met 195)	2002 B-1
58	hmd6h07	HUMRRL3A	ribosomal protein L3	52-180/880-(Met 7)	1256 B-1
59	hmd1b07	HUML6A	tumor antigen L6	583-1188(Met 109)	1188 B-2
60	hmd1c11	HUMCOX4A	cytochrome c oxidase (COX) subunit IV	150-152/152-(Met 35)	668 B-2
61	hmd1c12	HUMHISH3B	H3.3 histone, class B	304-352/352-(Met 109)	1039 B-2
62	hmd1d02	HUMIGF27	insulin-like growth factor II	833-1233/1233-(Met <1)	1621 B-2
63	hmd1f03	HUMECK	tyrosin kinase	3460-3921(Met 114)	3921 B-2
64	hmd1g11	HUMHSP90	90kD heat shock protein	2140-2205/2205-(Met 85)	2543 B-2
65	hmd2c04	HUMEF1AC	elongation factor 1-alpha	1442-1696(Met 16)	1696 B-2
66	hmd2c11	HUMEF1AC	elongation factor 1-alpha	1214-1696(Met 16)	1696 B-2
67	hmd2d03	HUMGFIBP	insulin-like growth factor (IGF) binding protein	848-852/852-(Met 124)	1462 B-2
68	hmd3b12	HUMGRP78	78 kdalton glucose-regulated protein (GRP78) gene	5104-5470(Met 593)	5470 B-2
69	hmd3c04	HUMEF1AC	elongation factor 1-alpha	1379-1696(Met 16)	1696 B-2
70	hmd3c12	HUMMTM1	mitochondrial cytochrome oxidase I and II URFs	2220-2768/2768-(Met 2212)	2771 B-2
71	hmd4f12	HUMPRPHOS1	phosphatase-1 catalytic subunit	1321-1367(Met 30)	1367 B-2
72	hmd5b05	HUMFBRAA	fibrinogen A-alpha-chain	1816-1888/1888-(Met 55)	2223 B-2
73	hmd5d07	HUMEF1AC	elongation factor 1-alpha	1152-1696(Met 16)	1696 B-2
74	hmd5e10	HUMFGFR4	FGFR-4 fibroblast growth factor receptor	2797-2915(Met 56)	2915 B-2
75	hmd5h09	HUMQM	Wilm's tumor-related protein (QM)	528-744(Met 42)	744 B-2

cDNA clone	Locus	Defintion	Match region and MboI site	Size	Fig.4 Class
76	hmd6a10	HUMMTM1	mitochondrial cytochrome oxidase I and II URFs	2288-2768/2768-(Met 2212)	2771 B-2
77	hmd6b02	HUMRPS17	ribosomal protein S17	319-477(Met 26)	477 B-2
78	hmd6h06	HUMGFIBP	insulin-like growth factor (IGF) binding protein	825-1462(Met 124)	1462 B-2
79	hmd2a08	HUMMNSODR	Mn superoxide dismutase	/767-829(Met 67)	829 C-1
80	hmd2e12	HUMNMTDC	NAD-dependent methylene tetrahydrofolate dehydrogenase cyclohydrolase	/1448-2102(Met 16)	2102 C-1
81	hmd2f05	HUMALB3	serum albumin	/1924-2055(Met 40)	2055 C-1
82	hmd5d05	HUME4BP4RN	E4BP4 protein	/1218-1923(Met 214)	1923 C-1
83	hmd5f08	HUMTLCA	ADP/ATP translocase, 3' end	/990-1116(Met <1)	1116 C-1
84	hmd5h02	HUMLLREP3	LLRep3	/536-934(Met 214)	934 C-1
85	hmd1e03	HUMODC	ornithine decarboxylase	330-841/(Met 88)	1812 C-2
86	hmd2b05	HUMRPSA3A	ribosomal protein S3a	1-66/(Met <1)	836 C-2
87	hmd2f12	HUMARF1A	ADP-ribosylation factor (ARF3)	19-180/(Met 425)	1058 C-2
88	hmd3d12	HUMENOA	alpha enolase	864-977/(Met 95)	1755 C-2
89	hmd1a12	HUMAGP1A	alpha-1-acid glycoprotein	16-81/(Met 16)	737 C-3
90	hmd3e02	HUMB61	B61 protein	21-131/(Met 74)	1480 C-3
91	hmd4f06	HUMEF1AR	elongation factor 1-alpha subunit	5-137/(Met 54)	1506 C-3

Table 2 Identified clones in 3'-directed cDNAs

cDNA clone	3'-directed cDNAs (library-frequency)	Match site	MboI site	Homology(5'-region) in PIR
hmd1c10	HUM00D0G12(HepG2-6/982)	(219/1)-54 (90%)	219	Rat Ribosomal protein L9
hmd1d06	mm0347(hl60-1/1268)	(251/1)-101 nt (81%)	251	
hmd2a07	pm0387(hl60d-3/1417)	(55/1)-285 nt (81%)	55	C.elegans Heat shock protein 70F
hmd2d01	pm0985(hl60d-2/1417)	(148/1)-331 nt (93%)	148	
hmd2f10	HUM00D3A07(HepG2-2/982)	(208/1)-106 nt (96%)	208	
hmd2h04	HUMHM02F11(HepG2-1/982)	(200/1)-50 nt (86%)	200	
hmd3a05	HUM00TW182(HepG2-1/982)	(14/1)-247 nt (95%)	14	
hmd3c05	mm0423(hl60-5/1268)	(38/1)-100 nt (98%)	38	
hmd3c06	pm1895(hl60d-1/1417)	(144/1)-141 nt (95%)	144	
hmd3d08	HUM000S331(HepG2-2/982)	(83/1)-89 nt (98%)	83	
hmd3g11	mm0384(hl60-1/1268)	(38/1)-100 nt (99%)	38	
hmd3h05	pm1783(hl60d-1/1417)	(437/1)-162 nt (82%)	437	
hmd3h11	mm08c06(hl60-1/1268)	(63/1)-98 nt (83%)	63	Mouse Transplantation antigen P198
hmd4c12	HUM000HC02(HepG2-1/982)	(212/1)-144 nt (92%)	212	
hmd5a09	HUMHM02A08(HepG2-1/982)	(72/1)-304 nt (81%)	72	Mouse MER5 protein
hmd5c12	HUM00D0C12(HepG2-2/982)	(192/1)-234 nt (90%)	192	
hmd5d04	HUMHM01C05(HepG2-1/982)	(219/1)-178 nt (83%)	219	
hmd6a04	mm3091(hl60-3/1268)	(239/1)-86 nt (95%)	239	
hmd6b03	HUM00D0D12(HepG2-3/982)	(127/1)-52 nt (96%)	127	
hmd6b09	pm2457(hl60d-1/1417)	(183/1)-41 nt (100%)	183	

Table 3 Homology to PIR with clones that have not been identified in GenBank

cDNA clone	Locus	Definition	Homology region	MboI site	ATG site	Frame
hmd1c10	R5RT9	Ribosomal protein L9 -Rat (192 a.a.)	(1/8)=(222/81)-340(100%)	222		1
hmd3c05	R5RT35	Ribosomal protein L35 -Rat (123 a.a.)	1-(38/46)=(152/84)-180(90%)	38	30	3
hmd4h10	S16868	Gene I3 protein -Mouse (74 a.a.)	(3/20)=(101/52)-416(87%)			3
hmd1a04	A29453	Phosphocarrier protein precursor -Bovine (362 a.a.)	(2/86)=(337/197)-517(87%)			2
hmd3c06	S03210	Heat shock protein 70F -C.elegans (487 a.a.)	(3/289)=(146/336)-283(81%)	147		3
hmd2c02	A34832	Tat-binding protein-1 -Human (404 a.a.)	(2/298)=(61/317)-178(80%)	61		2
hmd4c12	S10384	Transplantation antigen P198 -Mouse (203 a.a.)	1-(233/166)=(319/194)-384(75%)	232		2
hmd2c07	A39610	SLY1-20 protein -Yeast (666 a.a.)	(2/92)=(58/110)-411(73%)			2
hmd1d10	SYMTAT	Alanine-tRNA ligase -Silkworm (967 a.a.)	(2/193)=(83/219)-374(70%)	67		2
hmd5c12	JQ0064	MER5 protein -Mouse (257 a.a.)	1-(112/80)=(210/112), ,(211/199)=(310/232)-495(68%)	210	106	1
hmd3f06	S13370	EBER-associated protein -Human (128 a.a.)	1-(29/10)=(356/119)-511(68%)	367	17	2
hmd1b08	B41440	Protein disulfide-isomerase -Alfalfa (40 a.a.)	1-(110/21)=(169/40)-274(60%)	261		2
hmd1d01	A33879	Cytosol aminopeptidase precursor -Yeast (514 a.a.)	1-(82/333)=(165/360)392(60%)	165	79	1
hmd1g12	S08686	Finger protein ZFP-36 -Human (582 a.a.)	1-(98/24)=(235/69)-393(60%)		71	2
hmd5f08	COBYC1	Cyclin 1 -Yeast (546 a.a.)	1-(58/73)=(120/93)-235(57%)	155	94	1
hmd3a07	B39930	InternalinB -Listeria monocytogenes (248 a.a.)	1-(138/6)=(269/49)-485(52%)	269		3
hmd3e12	A39596	Progesterone receptor B form -Mouse (923 a.a.)	1-(8/278)=(70/298)-107(52%)	32		2
hmd6a02	A35363	Synapsin Ia -Human (705 a.a.)	1-(105/539)=(179/563)-217(52%)	16		3
hmd2d11	JX0145	Hemoprotein H-450 precursor -Rat (547 a.a.)	1-(46/1)=(156/37)-249(51%)	174	46	1

Table 4 Hydropathy index table (Kyte and Doolittle)

Amino acid		Table value	Amino acid		Table value
Ala	(A)	1.80	Met	(M)	1.90
Arg	(R)	-4.50	Phe	(F)	2.80
Asn	(N)	-3.50	Pro	(P)	-1.60
Asp	(D)	-3.50	Ser	(S)	-0.80
Cys	(C)	2.50	Thr	(T)	-0.70
Gln	(Q)	-3.50	Trp	(W)	-0.90
Glu	(E)	-3.50	Tyr	(Y)	-1.30
Gly	(G)	-0.40	Val	(V)	4.20
His	(H)	-3.20	Asx	(B)	0.00
Ile	(I)	4.50	Glx	(Z)	0.00
Leu	(L)	3.80	Xxx	(X)	0.00
Lys	(K)	-3.90	***	(*)	0.00

付録. シークエンスデータ


```
hmd3b11.seq
ctcagtgcccccttgggctecggtggtcttactctcctgcatgcagcagctgcagctgga
agaggctcagtggttcgtctgctgctggaagaggtgctgacccccactgtcagtgccagg
accttgacaccagaaaatggaggcacggcaggctacacggaaaaggagcagaaggcagc
ccggcggaacgggaggaacagcagcagaggcagcaggagcaggagagcgtgaacgaga
agagcagcgccgatctgcccctcagtgaccgagagaagagagctctggctgcagagcgc
cgaactcgtgccagctgggagccccctaaactctcaatccccctgactctgaaatcgtcaat
actcgnctgcttggagtgtggggcctcctccaaggcctgantccttttaaacct
ngactctctttctgnttcaaacgttgccttcaggatccttccgggaaatntttacag
gtnaaaatx
hmd3b12.seq
cgggagacagctagggctacatggcttaccagagcaattttatatggtactgcaacaaa
gcatttgacttataatgatctcatcaacaaggaaatgattctctctcaaacctcagacaa
tgaagactctgctagtgctgtaatatgtataactggactcaggaaactttgttaggaa
aaaattgaaaagaacttaagtctcgaatgtaattggaatctcactcagagtgaggatga
aactgctatagcctaaaggcctgtttactgctttccattagcagtgctcacatgcttt
gggtgggntntcgaagaaggaaatggccatcttaaaaaggcggtaaaaaaccctgggtta
gggtgtgtgttccacttcaaaaatgttctatataaccacctggggccatgtgcatctgg
ggngtggggnggttttctnccataaaggngccacex
hmd3c02.seq
tcaatacaaggaatatttattaaaaatagattatttagaggnggcacagtggtgtgat
ccctctccttactcaaatacagtttcaaaaaggaaactcatgagaaatttcaaaaaat
acaagtttttagatgtttatnctttgccttntttttaaagggttttcttctgtttgng
tcttcaactcgaanttagaantntgtaantaaagggttttnttttgncttanccca
ntttaaactcaatgtaaaaatgggtatataatgcatcagtacagattttcagctnttggca
acatctgctggax
hmd3c03.seq
caaccaaccagctctcactcaagaggccacactgcaacctcagcttaacatgaataac
aaagactggctcaggagcagggttgcaccagcctggggatcaccggaggctcagtagtt
caagaccagcctggccaacatggtgaaccccactctnctaaaaatgtgtalatctttg
tntgtctcctgttatgtgtgccaagggagtatttcaacaagcttcaaaaacagccnct
tantcagcgtggngcaaacagctgccattccngctctttatgcaagx
hmd3c04.seq
attttggtaagttgtttccattaaaaagttactgattttaaantcaataacttaaac
tgccacagcgaaaaaaagaaacaaagttggccacaaaaacatctccttctctgtngg
gtttacagatgcatgtttatcatfaaccngtcttttgcactnnccttaaatggcaatt
gnacaancagttctgagaccgttcttccaccnctggttangngtggggtggcngntt
nggntagtntctnttngcctctgngctttctgggx
hmd3c05.seq
ggccggcgcgcttggcggttgtgcagcaatggccaagatccgagctcgggaaatcca
ttgcccgtgtctcacagttattaccagactcagaagaaaacctcaggaaatctaca
agggcaagngtacaagccccctggcctcggcagtx
hmd3c06.seq
gggaagctgctgaaaaggctaaaatgtgaactctcctcatctgtgcagactgacatcaatt
tgccctatcttacaatggattctcctggacccaagcatttgaatatgaagttgacccgtg
ctcaattgaagggatgtcactgatccttctagttagcctagtactgctgactggcct
gtatgtacatggggctcttcaactgaggccttgcaagtcaagctggcctgtgcatgtttg
tagatggggcagaggaatctagaacantgggaacttagctattx
hmd3c07.seq
ggccttggtggaacatgtcaaggaggagtgtgaccgctggccccctggcaggcagacat
atgcaagacctatacagccagttctgaaatctctatccagatgatgatgeacatgca
acccaaggngatcagggaccctcccccttctctggcctctnagtgaacaaagcagca
aaacaaagcagttttatatagaaagattagaagcctggantaatcaggctttttaaata
tgtaatccccctgtaatagcatagggnttttggaaagcagctnctggngcttgggacat
cagtgggnccaaggttntctatccccctgggx
hmd3c08.seq
gatcataaagttctatcaaggagtctatcaaggcatccatgtoagtggtgctatgctgg
ttacaacttgagatttttgaataaaaaatgtgcatatccatgcctctaaa
hmd3c10.seq
ccaagctcagtaacttcgggatgaagggaggctctgcttgcacaaacagagactcaag
tgtgccagcttccaaaaatttggagannagccttcaaaagcattgggcagtagctcgcctg
agccagagatttccaaaagctgagtttgcaagatttccaagttagtgacagatcagggg
agaccctaactgccagntatattnaatgtacaaantgaaaccagntgnaatagtgctc
tgtcagctgnantnttaangntatagtatattataactctx
```



```

::::::::::::
hmd5h11.seq
::::::::::::
ctttaaactactcagaggacaggatctctactaccacaaggaaaatagtttagggaaaac
cagcttttactgttttgaaaaattacagcttcaccctgtcaagttaacaaggaatgct
gtccaataaaaagtttctccaaa
::::::::::::
hmd6a02.seq
::::::::::::
gtttatcggngagctgatcagatatacganaacgccatgatgctgctggacttgttgac
gacctagggccatggggccgcttgaatganctgctttnaaggccctggagcgacac
tnacagccagggggccagaaggactgacaccacagatgacagccccctcttnagctt
tattaccataaattaaaggtatttcttaaccggaaa
::::::::::::
hmd6a04.seq
::::::::::::
cagcgcagggccgggaagcttctcgnagccctagcgtcgggccccgcttctcggggccg
ggtcccaggcagttcaattaacctccagaagatggctgaacctgcaggaataccagagca
agaaactgatgtctgcacaacggagtangagttcaaaagattctttagcagacactgcaa
atgaagctctcagggctgctaaagactaaatgcaaaagaaatgttttaaagcccaaga
tctatgtggtaaaaatgcacaggagcttggtagactcggggaaagagagagctcttctc
gnatgtttaccagctcgtggtgntataaccccgtaggtgtattctttaaattccca
gcccttx
::::::::::::
hmd6a05.seq
::::::::::::
gttetaatcttctctctctcttgataaacctccacagcttctgatttggcttctgcttc
tgcaactacccttagcatattcaccaggattcttctctctctctctctctctctctctc
tctcttctctctctctctctctctctctctctctctctctctctctctctctctctctc
tgccacgatctgtgagaaaataaattctgttgttaagccaaa
::::::::::::
hmd6a06.seq
::::::::::::
gcaagcagctgccagggtcatcgcattggccaaaacaagtctgttaaagttatcggct
gattggctcgagacactgtggaagaaatagctctataggaaagcagctccaaactgcagct
caccacatgatctggggccctcagaattnctctctctctctctctctctctctctctctc
ctggatgttttgctctgttctgtacctgtaataatagggaacaacaacttttttggga
aagcccttgacccagcttgctagtgcataataataaattctctctctctctctctctctc
::::::::::::
hmd6a08.seq
::::::::::::
ggaecgnnnntncaacctggatgaaaatncaaaaatggngtaatcagtaatacaaggn
tatttcancaagggtgatgggtattgaagttagccagaataggagcagcaggagct
tctcttctctctctctctctctctctctctctctctctctctctctctctctctctctc
gtcttctgtgagatcagtcataaataantctctctctctctctctctctctctctctctc
tattttncaatgttctctctctctctctctctctctctctctctctctctctctctctc
tacaatnctnngttaaantaggggcagatcctctcncacgggttctngtaantgttnag
gggtaax
::::::::::::
hmd6a10.seq
::::::::::::
ggaecgnnantagcctggagctctaggcagcagctcgnngntctctctctctctctctc
gggcccagccagcaaccttntaggtaacgaccacatctacaagttatctctcagccca
tgcantntaataatctctctctctctctctctctctctctctctctctctctctctctc
actagtctccctaaataatcgggtgccccgataatgggcttctcccgataaacaataag
ctctctnctctnctctctctctctctctctctctctctctctctctctctctctctctc
agcaggnacaggttgaacagntactcctctctctctctctctctctctctctctctctc
tncgtagncttaccatcttctctctctctctctctctctctctctctctctctctctc
anatttgatnaaaaataatctctctctctctctctctctctctctctctctctctctc
::::::::::::
hmd6a11.seq
::::::::::::
catttatgagaggatggttccaccagggcagacagcctgacaggttacggagttccta
tgctagtcagcatagtcagcttgggcaagacctctctctctctctctctctctctctctc
cattactctctctctctctctctctctctctctctctctctctctctctctctctctc
tggaaactgtggagctccaaggatctaagatggccggcgggctgggctcccgctgtaac
ccagcaacttgggagggcggagggcggcgnatgcgttgaggctcngggagttngagctccag
tctgtactaaaaatacaaaaataggccanacaaggnctggngcatgctataaatctccagg
tantnaggagx
::::::::::::
hmd6a12.seq
::::::::::::
cacagggtctgtgaaaactaccctaaaagcaaaaatgggaaaggaaaagactcatatca
acattgtctgcatggacacgtagattcgggcaagtcaccactactggccatctgatcg
ccgttctggtaaaaagctggaagatggccctaaattcttgaagctcgggtgatgctgccat
tgtgataatggttctcggcaagccatgtgtgtgagagctctcagactatccaccttct
gggtcgttctgtctctgtgataagagacagacagttcgggtgggtgtcatcaaagcagt
ggacaagaaaggctctctggagctggaagggtcacaaggtcttggccagaaaagctcnc
agaaggcttaaatggattatnctccctaaataccctgex
::::::::::::
hmd6b01.seq
::::::::::::
agaaaaatagtttagaggaagaggactgctgtccaggctctgttttgaagcacagagctc
tctccaactctcaccatgtttagcagcatttagggcaattatcaactcaaaaattcca
caaatggagaataatctcaaaaggtaattcttaattctcaaaaggctatcgaaaga
aatgacaagcaacttgcctgactgctcaaaaatttggcgcactcttataaccaaaat
tgggcaacagaaagttttagcttagagatgctcacttctcctggaaagnccaacagat
gatagcgttgaagtagctattggtttctttaaaggaatgtggcctcaaaaatttaaccacc
aaagggtgcacccaaggagggatcaatgctatatttganccgctctcgaacaattctg
aattgagctngaantgcccnaaagggtcex

```



```

: : : : : : : : : :
hmd6b02.seq
: : : : : : : : : :
caagttagtgacagatcctgacactaaggaaaagctggaagctttggacttggcagctt
gtccaaacctcaggtcactcagcctacagttgggatgaatttcaaaacgctcggggacc
tgtttgaatttttctgtagtgctgtattattttcaataaaatctgggacacacaaaaata
aa
: : : : : : : : : :
hmd6b03.seq
: : : : : : : : : :
gtgcttagcatccgtncatgaaacaaaatnatgtgatacgataaaataggatgtnatgta
taaatnataaatagtagcagtggaagcactcctactggaagaccatactnactctgtna
catctccaaggcagttcttggagttggccaggagatcacgacctgcaactccagcct
ggcgacagagcgagactccatctcaaa
: : : : : : : : : :
hmd6b05.seq
: : : : : : : : : :
aatgaagtgggtaacctttnatttccctcttttctcttttagctcggcttattccagggg
tggtttcgtcagatgcacacaagagtgaggttgcctcctcggtttaaaagattgggaga
agaaaaattcaaaccttgggttgattgctcttctcagtagtctcagcagtgctcatt
tgaagatcagaagaaacctttttagcaaatatcaagaatctagtttcaactcagaaactt
ctgattagctctcagctctcagctctatttctcttaggagtttaattgacctctctcc
ccatcttccctcaagtccataaaaaccttcaattacacataaagtagcagtggggctcag
tgaatctgcttgccttctgaaaagttctgggcttaagattccagctctggttcaatca
ntataggcacctgtccaaa
: : : : : : : : : :
hmd6b06.seq
: : : : : : : : : :
gtcagcagcggatgccccccaggtaagaaagccgaaccagagccgatgagagaggag
gagaaaaataatctcctcagaaacctgaaatccagccaaggctccaagtagctgagtg
natgctgtccccagcagcagatcggcctggagccagtgccaagtagcggtttggtgtg
tacctgaaaaacaaactctctcagctctctggggcggaggtttctgctggttttcttct
ctcaaaagaaaaaacatgcaccattttcaatgtgaaa
: : : : : : : : : :
hmd6b07.seq
: : : : : : : : : :
aagaacgcccgtctctcgcctctcagcgcggcttctccttctgtnccggagcgcctctct
cagcctcgggctctctnggctcgtctgctcctcgcagcctcaccggctcagagacc
atggccgagcggggaaactcgaacttgaccggcgaacacagaacacagagagtggtgta
gcaaggttccataaatattgtnacagcaatgggctaaagctctggaagaggtgaaagtg
ggaaactcgggatgccaagactcaaggaggtctnagggtgctattactttgaaatgag
gatcacgtttcgtgtctactcaaacggttaataaaaaggcgagagtttctcaagtgn
aaa
: : : : : : : : : :
hmd6b09.seq
: : : : : : : : : :
tatcactattacataacatactcaggcaaaagaactttgctcagggacataccatgtaa
tattttttgtttctttacagactagtctacagctcctgcttactcaaaacaaacaaat
aaactataaccttataataagtattatgtagttagtaactacctcctnagtttgaca
cagatctcgttttgacaataaaaatccttactactttctcaaa
: : : : : : : : : :
hmd6b10.seq
: : : : : : : : : :
gcccgaaccggcggcgggacaggttgctgcttctcctcctcggcctcaccattccaga
ccaaaattgaaaaaatgggtgacctcaccaggttaaggatgagaaagatttctaggtt
ttncatcctatgacaacaaatnatncttcaaaaatnatgcttatgagtagcgaactgcat
tctatagattgacaagaaaggttttggccaatccagaagactgtgtagcatttggcaaaag
gagaaaaatgccaagaagtagcttctgaaacagatgacagagtagaacgttagcagagccc
acctgaattgacctgaaaaataattatccaatttctttggaattggcctcctgtattc
cttnnagtggtcccggancccttanaagccatnctgaaactcagccctatttx
: : : : : : : : : :
hmd6b11.seq
: : : : : : : : : :
gagaatttccattaggcaagctgtatgacagtggaatncccttttaggagacacacaaat
aagactctctggtctctnctcttggcttgggagactttagacatttccaaagcacaaaata
aaagcctcgtatttccccattgagagttttgtnccaaggaatagagtgagacatattgg
gtgagtcataataatacaaaataatttatgaagagctgggtctgcaatagctagctaaaa
nctacttggtgtcagctctctggttatagtagtataaagagcctgaggaggtctggcaaga
tagatggtgtattatttgatcctgagagattttccatctctgtcaccattcacttg
caattgaaaaagnttcccttgcctgttggtggcctagntcctcttgnaccatatttac
ntaatgngggttaaax
: : : : : : : : : :
hmd6c01.seq
: : : : : : : : : :
gtgaaatcagcctctcgtcagctcaagctgagcttagcgcgggtggagcaggggtcactg
agtcagcatctggatcaaggaataatataatcttctcaaaacttataaatctggtg
gacaaaataatcatttctccttggcagcaaaaagttctcagtagcatttttggtggtt
ctctttgaaaagaaaagctgaaatattataaatgctagtagttctgcccattatga
aagatgaaataaagtagtcaaaataataacatttcaataataaaggtgattattga
gaagtaaggtttgaagggctataaagaaaactgtttatccctgggtaattatatta
nggggatttgctnctgggtncaataax

```



```

::::::::::::
hmd6c02.seq
::::::::::::
cgcaagatggcagccaccacgggctcggacaatttatgaaaaccgaatatacagccttaa
aatagaatgtggacctaaataccagaagcaccoccttfgtaagattgttaacaaaaat
taatatgagatggagtaaatagttctaatggagtggtggaccaagagccatcagtgcc
tagcaaaatggcagaattcatatagcatcaaatgttctctgcaagagctcggcgctaa
tgatgtctaaagaaaatagaaactccctcagcgcgccccgaggncaagtgttacagcaat
taaatcanaaggnnaancncagggccctccccctccccccnatttcggtttaatean
gtcttcanttccccagggtggtaaattttncttggtacggcnttntaggcctnaaaagg
ccccggx
::::::::::::
hmd6c03.seq
::::::::::::
ctaattattactcagaaaaaggatttcttaagagttgctgtaaacctgaagattgtactc
cacagagagatgcagacaaagtaacaatgaaggttgtttataaagggtgatgactctcg
tttttttggtcgaaactggtcgctgccatcgtaggattgttttcagacatgagattaaga
acagctccaactgattggaatctttctcgcctactgcctctctcgtgccataaaaaa
accagtatgagatagtgtaaccaatgtatctgtgggctatctctctaccttaaggg
catttagggctccccctggggattagaagttgcttggctggngganctgncaacaacta
cctanctgtaggccaaaaanctacaccagagntggttcaatcanggtx
::::::::::::
hmd6c04.seq
::::::::::::
ggcgtcaacaagaatcactcggagggcgccggagtgatctgtttagtgagatttagcat
gtgtgaataaagtatatgcaggaggaaa
::::::::::::
hmd6f08.seq
::::::::::::
ggaaccgaggagtcggacgcaggcagaccatgtggaccctggtgagctgggtggccttaa
cagcagggctggtggctggaacgcggtgccagatggtcagttctgcctctggtccctgct
gctggaccgccggaggagccagctacagctgctgcgctccctctcggacaaatggcca
caacactgagcaggcatctgggtggccctgcaggttgatgccactgctctcggccc
actcctgcatctttaccgtctcagggacttccagttgctgcccctccagagggcctgggn
atgcggggatgggatantgttgccaaggggttccaatnaagtanacggnggtttt
anatagntttggaaaatnttttccnctgnntttggtttttgggaaaaaattttggg
ccaagntttgggaaaatttggccaagntttggattttggaaaanagcttatggx
::::::::::::
hmd6h06.seq
::::::::::::
ctggctcgaactcctgactctgtgatcaggggagacccaactgccagatataatataa
tgtacaaaactgaaaccagatgaaataatgttctgtcacgtgaaatatttaagtatatag
tatattatactctagaacatgcacatttatatatatgtatagtatatatatagn
anctactttttatactccatacataacttgatagaaaagctgtttatttctactgta
agtttatttttctacacagtaaaaacttgtactatgttaataacttgtcctatgtcaat
ttggtatcatgaaaccaactctcatcatatggatggaaaggaattgcctttctgnc
ccttcccaaggtcccgggectgtttttaangggctngggaaaatnctgcccggnaaat
gnaaaatggattaggggggttggttx
::::::::::::
hmd6h07.seq
::::::::::::
gatggngtgatgtttcacagaaagtctccgatcccagacatgggtccctcggttctn
cctcggaaagcganagcagccatcgtgggaaggtgaagagcttccctaaggatgaccat
ccaagcggctccacctcacagccttctctgggatacaaggctggcatgactcacatcgtgc
gggaagtgcagagccgggatcaagaacaatgcctccactgactatnacatctcgacaa
gagcatcaacctctgggtggctttgtccactatggtgaagtgaccaatgacttntcat
gctgaaaggcttgtgtggtgggaacaagaagcgggtgnacccttccggaaagtcttnc
ttgtgx

```




inches 1 2 3 4 5 6 7 8
cm 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19

Kodak Color Control Patches

© Kodak, 2007 TM: Kodak



Kodak Gray Scale



© Kodak, 2007 TM: Kodak

A 1 2 3 4 5 6 **M** 8 9 10 11 12 13 14 15 **B** 17 18 19

