

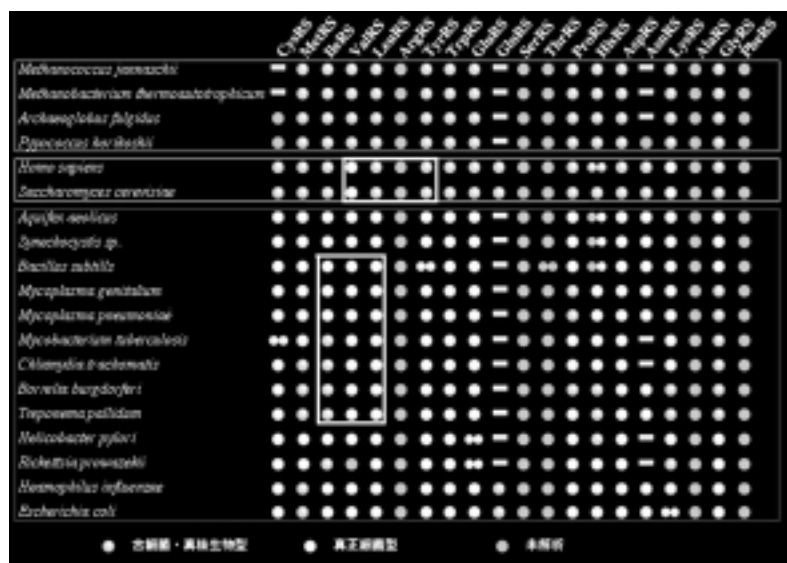
プロテオーム配列空間からの正規構造の抽出

(財) 癌研究会癌研究所 芝 清隆

【序論】プロテオームをざっと眺めてみると、いろいろな階層の「繰り返し」構造が目につく。一次構造モチーフの繰り返し、立体構造レベルでの繰り返し、モジュールの繰り返しなど、「繰り返し性」が生体高分子の構築原理として重要な役割を果たしていることが伺われる。本研究では、このようなプロテオーム正規構造としての繰り返し性の研究の一環として、ペプチドの一次構造レベルの繰り返し性を、進化的に関連のあるタンパク質ファミリーに焦点を絞り解析するシステムを構築した。タンパク質ファミリーが形成するプロテオームを「パラログとしての縦軸」「生物種としての横軸」の2つの座標軸から亜群に分割し、これらの亜群内でのペプチド繰り返しの特徴を抽出し、繰り返し性の起源を求める研究である。

【方法】webを通して、広く利用してもらうことを目的に、JAVAによるプログラムの構築を進めた。癌研内に設置したサーバーにデータベースを構築し、使用者はこのデータ群から任意に亜群を作成し、その中の繰り返し配列を抽出することが可能である。

【結果】今回は、タンパク質ファミリーとして、この種の解析には頻用されるタンパク質ファミリーの1つである「アミノアシルtRNA合成酵素」ファミリーを選んだ。この酵素ファミリーは、酵素としては最も起源の古い酵素であり、全ての生物が、基本的には20種のアミノアシルtRNA合成酵素を利用することによって、遺伝情報の翻訳処理をおこなっている。20種の酵素は、少なくとも2つの起源の異なる祖先遺伝子から派生してきていることが分かっており、20種の酵素の中に、分散進化と収斂進化のどちらのタイプの進化様式も見出すことが可能である。さらには、アミノアシルtRNA合成酵素遺伝子の一部が、生物種を飛び越えて、ゲノム間を移動していることも明らかになっている。立体構造の解析も急速に進んでおり、タンパク質の構造・機能・進化を研究する材料としては、他に類を見ない好条件を備えていると言える。既にゲノムの全構造が解明されている22種の生物がもつアミノアシルtRNA合成酵素配列のデータベースを構築し、その中から自由に亜群を作成できるようにした。例えば、「古細菌のメチオニルtRNA合成酵素とその兄弟であるイソロイシルtRNA合成酵素のデータセット」「大腸菌がもつロスマンフォールド型の10種のアミノアシルtRNA合成酵素のデータセット」「全てのアルギニルtRNA合成酵素のデータセット」などの選択が可能である。さらには、酵素を進化的に古いと思われる「活性部位」、比較的新しい「第2ドメイン」、高等真核生物に特有の「付加ドメイン」の3つの領域に分割し、これらの領域を含めた、より複雑なデータセットの構築もできるようにしてある。



このように作成したアミノアシル tRNA 合成酵素亜群のデータセットの中から、「任意の長さ以上」のペプチド配列を「任意の回数（以上）」、1つの読み枠の中にもつものを「repetition」コマンドを用いて、抽出して行うことができるようになった。

【展望】完成したシステムを利用することから、特定のペプチド配列の遍在性を、「生物種」「パラログ内」「ドメイン別」などのいろいろな角度から抽出して調べることが可能となった。このような解析を進めることにより、繰り返し構造の進化的起源を類推することが可能である。構築したシステムは、他のタンパク質ファミリーにも展開可能で、広くプロテオーム研究に利用されるものと期待している。