

塩基配列からアミノ酸配列への翻訳

塩基配列をアミノ酸に翻訳するには、その塩基配列が何の配列で、どのような遺伝暗号によってアミノ酸をコードしているのかが分かっている必要があります。ここではすでにそれは分かっているものとして、塩基配列をアミノ酸配列に変換する方法を説明します。

MEGA を使う場合

Windows 環境なら MEGA の AlnExplorer を使うのが簡単です。FASTA 形式でデータを用意して MEGA で開くと、AlnExplorer が起動するので「Data」メニューの「Select Genetic Code Table」から遺伝暗号表を指定します。その上で、「Data」メニューの「Translate/Untranslate」を使えば翻訳は終わりです。ただし、MEGA の場合、塩基配列に縮重コード（「A or C」を「M」と表記するもの）があると、本来は何らかのアミノ酸「X」に翻訳すべきなのにギャップ「-」としてしまいます。アミノ酸配列の用途によっては問題が起きる可能性があります。翻訳後の配列は「Data」メニューの「Export」からファイルに書き出すことができます。

EMBOSS を使う場合

EMBOSS の transeq コマンドにより FASTA 形式の塩基配列をアミノ酸配列に翻訳することができます。MEGA のような問題はありませんが、コマンドラインから操作する必要があります。Jembooss などの GUI を介した操作もできますが私は利用していません。他にも EBI が Web インターフェイス を用意していますのでこちらの Web フォームから翻訳結果を受け取ることもできます。

transeq での翻訳には、まず遺伝暗号表の番号を知る必要があります。マニュアルのページが参考になります。Command line arguments の `-table` の項を参照して下さい。これだけで分からないなら NCBI のページも見ましょう。

番号が分かったら、以下のようにコマンドを実行します。

```
transeq -table x -frame x input-file-name output-file-name
```

入力配列が第 1 コドン位置から始まっていない場合、`-frame` オプションに 1 ~ 3 を指定することでずらすことができます。F を指定すると 1 ~ 3 の全パターンでの翻訳を行います。-1 ~ -3 を指定すると逆方向に翻訳します。R を指定すると -1 ~ -3 の全パターンでの翻訳です。6 なら 1 ~ 3 と -1 ~ -3 の全パターンでの翻訳になります。これを利用して最も終止コドンの出現が少ない翻訳パターンを探すことで正しいアミノ酸配列と塩基配列のコドン位置を推測できます。ただ、そういう目的の場合は transeq を使うよりも showseq コマンドを使う方が良いでしょう。以下のようにコマンドを実行すると、見やすいテキストファイルが作成されます。ただのプレーンテキストなので適当なエディタやページャで中を見て下さい。

```
showseq -table x -format 6 input-file-name output-file-name
```

ファイルに書き出さずにページャ (less) で表示するだけなら以下のようにします。

```
showseq -table x -format 6 input-file-name stdout | less
```

transeq -frame 6 と showseq -format 6 をいっぺんにやってしまう sixpack というコマンドもあります。以下のように使います。

```
sixpack -table x input-file-name
```