W3.1: アラインメントの場合の数

以下の 2 つのアミノ酸配列のアラインメントを考える。

KSALTIQLI
KSTLANIFEQV

例えば、以下のアラインメントの配列類似度および配列一致度は

```
K S A L T I Q L I - -
K S T L A N I F E Q V
+5+4-1+4-1-3-3+0-3-8-8        配列類似度 -14  配列一致度  27.3%
K S A L - T I - - Q L I
K S T L A N I F E Q V -
+5+4-1+4-8+0+4-8-8+5+1-8    配列類似度 -4   配列一致度  41.7%
K S A L - T I - Q L I
K S T L A N I F E Q V
+5+4-1+4-8+0+4-8+2-2+3       配列類似度 +2   配列一致度  36.4%
```

このように、アラインメントのとり方は何通りもあり、それぞれ、配列類似度と配列一致度が異なる。
一般に、長さ m と長さ n の 2 つの配列のアラインメントの数は以下の式で与えられる。

$$f(m,n) = \sum_{k=0}^{\min\{m,n\}} 2^k {}_mC_k \, {}_nC_k = \sum_{k=0}^{\min\{m,n\}} \frac{(m+n-k)!}{k!\,(m-k)!\,(n-k)!}$$

これは、ギャップと対応する部分の文字を k として、どちらの配列にギャップが対応づけられるかで2^k通り、さらに k を m, n からどう選ぶかで${}_mC_k \, {}_nC_k$通りが考えられる。
例えば、具体的に計算すると、以下のように膨大な数になることがわかる。

$f(1,2) = f(2,1) = 5$

$f(4,2) = f(2,4) = 41$

$f(8,4) = f(4,8) = 3649$

$f(16,8) = f(8,16) = 39490049$

$f(100,100) \approx 2 \times 10^{74}$