1 Globálny a lokálny alignment sekvencií

Implementujte Needleman-Wunshov algoritmus na globálny alignment s lineárnymi pokutami za diery. Vaša funkcia bude volaná nasledovne

```
NeedlemanWunsh(seqA, seqB, match, mismatch, gap)
```

kde

`seqA` je string obsahujúci sekvenciu,

`seqB` je string obsahujúci sekvenciu,

`match` je kladné skóre za zhodu,

`mismatch` je záporná pokuta za nezhodu,

`gap` je záporná pokuta za dieru dĺžky 1.

Implementujte Smith-Watermanov algoritmus na lokálny alignment s lineárnymi pokutami za diery. Vaša funkcia bude volaná nasledovne

```
SmithWatermann(seqA, seqB, match, mismatch, gap)
```

kde

`seqA` je string obsahujúci sekvenciu,

`seqB` je string obsahujúci sekvenciu,

`match` je kladné skóre za zhodu,

`mismatch` je záporná pokuta za nezhodu,

`gap` je záporná pokuta za dieru dĺžky 1.

Pomocou oboch funkcií s parametrami `match = 1`, `mismatch = -1` a `gap = -2` spočítajte skóre pre všetky dvojice sekvencií

1. http://ksvi.mff.cuni.cz/~mraz/bioinf/C.dna,
2. http://ksvi.mff.cuni.cz/~mraz/bioinf/F.dna,
3. http://ksvi.mff.cuni.cz/~mraz/bioinf/G.dna

Výsledné skóre porovnajte.

2 Needleman-Wunschov algoritmus s afinným skóre za diery

Implementujte Needleman-Wunshov algoritmus na globálny alignment s afinnými pokutami za diery. Vaša funkcia bude volaná nasledovne

```
NWAffine(seqA, seqB, match, mismatch, gapopen, gapext)
```

kde

`seqA` je string obsahujúci sekvenciu,

`seqB` je string obsahujúci sekvenciu,

`match` je kladné skóre za zhodu,

`mismatch` je záporná pokuta za nezhodu,

`gapopen` je záporná pokuta za otvorenie diery,

`gapext` je záporná pokuta za predĺženie diery o jeden znak.

Volaním tejto funkcie s parametrami `match = 1`, `mismatch = -1`, `gapopen = -2` a `gapext = -1` spočítajte skóre pre všetky dvojice sekvencií

1. Homo sapiens insulin (INS),
2. Rattus norvegicus insulin 1 (Ins1),
3. Mus musculus insulin I (Ins1).

Uvedené sekvencie môžete získať z "NCBI Entrez sequence retrieval system" na adrese http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?CMD=search&DB=nucleotide.

Hľadajte (iba v sekcii nukleotidových sekvencií) dotazom tvaru:

```
human[Organism] AND insulin[title]
```

3 Hľadanie v bioinformatických databázach

V knihe

M. Zvelebil, J. Baum (2008) Understanding bioinformatics, Garland Science, Taylor & Francis Group, LLC, ISBN: 0815340249 ISBN: 9780815340249
http://www.garlandscience.com/product/isbn/9780815340249?fromSearchResults=fromAlphaSearchResults
je uvedený obrázok Obr. 1 porovnávajúci lokálny a globálny alignment. Zistite, ktoré sekvencie sú tam porovnávané. Keď budete mať sekvencie, tak urobte globálne a lokálne alignmenty pomocou nástrojov z European Bioinformatics Institute (EBI) http://www.ebi.ac.uk. Výsledky porovnajte.

Pre jednoduchosť nasledujú opísané postupnosti báz proteínov z obrázku

PI3-kinase:

```
HQLGNLRLEECRIMSSAKRPLW
LNWENPDIMSELLFQNNEIIFKNGDDLRQDMLTLQIIRIMENIWQNQGLDLRMLPYGCLS
IGDCVGLIEVVRNSHTIMQIQCKGGLKGALQFNSHTLHQWLKDKNKGEIYDAAIDLFTRS
CAGYCVATFILGIGDRHNSNIMVKDDGQLFHIDFGHFLDHKKKKFGYKRERVPFVLTQDF
LIVISKGAQECTKTREFERFQEMCYKAYLAIRQHANLFINLFSMMLGSGMPELQSFDDIA
YIRKTLALDKTEQEALEYFMKQMNDAHHGGWTTKMDWIFHTIKQHALN
```

cAMP PK:

```
GNAAAAKKGSEQESVKEFLAKAKEDFLKKWENPAQNTAHLDQFERIKTLGTGSFGRVML
VKHMETGNHYAMKILDKQKVVKLKQIEHTLNEKRILQAVNFPFLVKLEFSFKDNSNLYMV
MEYVPGGEMFSHLRRIGRFSEPHARFYAAQIVLTFEYLHSLDLIYRDLKPENLLIDQQGY
IQVTDFGFAKRVKGRTWTLCGTPEYLAPEIILSKGYNKAVDWWALGVLIYEMAAGYPPFF
ADQPIQIYEKIVSGKVRFPSHFSSDLKDLLRNLLQVDLTKRFGNLKNGVNDIKNHKWFAT
TDWIAIYQRKVEAPFIPKFKGPGDTSNFDDYEEEEIRVSINEKCGKEFSEF
```

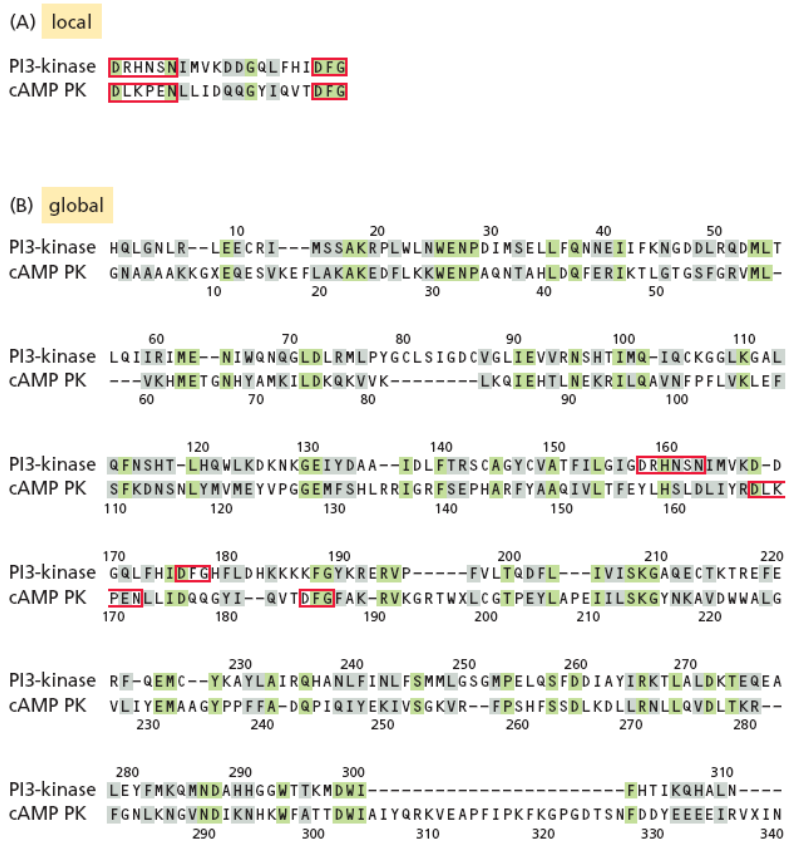

Figure 4.7
Local and global alignments. The complete sequences of PI3-kinase p110α and the cAMP-dependent protein kinase (cAMP PK) shown in Figure 4.5 were compared. (A) Local alignment using the program LALIGN (a subset of the FASTA package) has matched a short conserved region in the kinase domains that contains the functionally important residues D and N in the DLKPEN sequence and the DFG repeat common to nearly all kinases. (B) Because of the low overall sequence similarity, a standard global alignment of these two sequences using the program ClustalW has not matched these functionally important residues (boxed in each sequence). Green shading, identical amino acids; gray shading, similar amino acids.

Obr. 1: Obrázok z knihy