

## Programowanie dynamiczne – algorytm

### Zad. 1

Wykonaj przyrównanie globalne dwóch sekwencji DNA (na [wydruku](#)) przy pomocy algorytmu Needlemana-Wunscha.

Przyjmij poniższy system punktacji:

- dopasowanie/zgodność (*match*): +3
- niedopasowanie/niezgodność (*mismatch*): -1
- kara za przerwę (*gap penalty*): -2

W odpowiedzi podaj uzyskane dopasowania oraz wartości ich punktacji.

### Zad. 2

Wykonaj przyrównanie lokalne sekwencji z poprzedniego zadania (ten sam system punktacji) przy pomocy algorytmu Smitha-Watermana. W odpowiedzi podaj uzyskane dopasowanie oraz wartość jego punktacji.

## Przyrównanie sekwencji - podstawowe informacje

### Zad. 3

Poniżej znajduje się przyrównanie dwóch sekwencji DNA w dwóch formach zapisu. Odpowiedz na pytania.

dna1	GA-TACTA-	dna1	GA-TACTA-
	.	dna2	GATTTC-AA
dna2	GATTTC-AA		

1. Ile wynosi długość przyrównania?
2. Na ilu pozycjach przyrównania nukleotydy są dopasowane (*match*)?
3. Ile wynosi procent identyczności przyrównanych sekwencji?
4. Ile wynosi wartość punktacji (*score*) przyrównania sekwencji? Przyjmij poniższy system punktacji:
  - *match*: 2
  - *mismatch*: -1
  - *gap*: -2

### Zad. 4

Poniżej znajduje się przyrównanie dwóch sekwencji białkowych w dwóch formach zapisu. Odpowiedz na pytania.

s1	MSSEATGEW	s1	MSSEATGEW
	. : :	s2	MSKQISG-W
s2	MSKQISG-W		

1. Ile wynosi procent identyczności przyrównywanych sekwencji?
2. Punktacja dopasowania/niedopasowania każdego aminokwasu zależy od użytej macierzy substytucji. Najczęściej używa się macierzy BLOSUM62 (<https://ftp.ncbi.nlm.nih.gov/blast/matrices/BLOSUM62>). Ile wynosi wartość punktacji w tej macierzy dla:
  - dopasowania M-M
  - niedopasowania E:Q
  - niedopasowania S.K
  - niedopasowania A.I?
3. Ile wynosi procent podobieństwa przyrównywanych sekwencji?  
*Procent podobieństwa* sekwencji aminokwasowych jest zawsze większy lub równy od procentu identyczności. Podobieństwo, oprócz identycznych reszt aminokwasowych, uwzględnia również substytucje aminokwasów podobnych (tj. takie substytucje aminokwasów, które są dodatnio punktowane w użytej macierzy substytucji, w tym przypadku macierzy BLOSUM62). W przypadku sekwencji nukleotydowych podobieństwo jest tym samym co identyczność.
4. Podaj wartość *score* całego przyrównania przy zastosowaniu macierzy BLOSUM62 i kary za przerwę: -10.

## Programy on-line: Needle i Water

### Zad. 5

W pliku <http://combio.pl/files/insulin.fasta> znajduje się sekwencja genomowa genu insuliny oraz sekwencja kodująca tego genu (CDS, *coding sequence*). Korzystając z internetowej wersji programów needle i water (<http://www.ebi.ac.uk/Tools/psa/>) wykonaj - w dwóch kartach przeglądarki - przyrównanie tych sekwencji. Ustaw typ porównywanych sekwencji (*Enter a pair of*) jako DNA. Umieść pierwszą sekwencję w formacie FASTA w pierwszym oknie i drugą sekwencję w drugim oknie. Wykonaj przyrównanie.

1. Z ilu egzonów składa się badany gen insuliny?
2. Podaj pozycję początku i końca egzonów w sekwencji genomowej.
3. Dlaczego procent identyczności przyrównania w programie needle jest niższy niż w water?

#### Zad. 6

W bazie UniProt znajdź rekord acylfosfatazy u: człowieka (ID: P14621), *Escherichia coli* (ID: P0AB65) oraz *Salmonella sp.* (ID: A0A3D1C3E6). Wykorzystując program needle wykonaj przyrównanie następujących par sekwencji:

- *Escherichia coli* : *Salmonella sp.*
- *Escherichia coli* : człowiek

1. Które przyrównanie odznacza się wyższym poziomem podobieństwa sekwencji i ile wynosi?
2. Czy wynik przyrównań jest zgodny z Twoimi oczekiwaniami na temat mutacji sekwencji?
3. Podaj kilkuaminokwasowy fragment sekwencji najbardziej zachowany w trzech analizowanych sekwencjach.  
- Jaka może być przyczyna zachowania tego fragmentu?
4. Jaka macierz substytucji aminokwasowych została użyta w tych przyrównaniach?

#### Programy lokalne: needle i water

#### Zad. 7

W dwóch plikach znajdują się dwie sekwencje proteazy serynowych: z *Bacillus lentus*

([http://www.combio.pl/files/protease\\_bacillus.fasta](http://www.combio.pl/files/protease_bacillus.fasta)) i człowieka ([http://www.combio.pl/files/protease\\_human.fasta](http://www.combio.pl/files/protease_human.fasta)). Pobierz oba pliki i wykonaj przyrównanie tych sekwencji używając lokalnie zainstalowanego programu water. Wyświetl pomoc tego programu (np. `water -h`) i uruchom go, aby przekazać wszystkie argumenty w jednym poleceniu.

1. Ponownie przyrównaj te sekwencje programem water, tym razem zmniejszając karę za otwarcie przerwy = 1. W jaki sposób zmniejszenie kar za otwarcie przerw wpłynęło na to dopasowanie?
2. Wykonaj przyrównanie jak w poprzednim punkcie zwiększając kary za stosowanie przerw: otwarcie przerwy = 25, wydłużenie przerwy = `5`. W jaki sposób zwiększenie kar za stosowanie przerw wpłynęło na to dopasowanie?
3. Jak nazywa się model karania za przerwy przyjmujący dwa parametry: otwarcie i wydłużenie przerwy? Wskazówka: [https://en.wikipedia.org/wiki/Gap\\_penalty#Types](https://en.wikipedia.org/wiki/Gap_penalty#Types).
4. Jak nazywa się model karania za przerwy, który używałeś/aś w zadaniach 1-3?

#### Zad. 8

Wykonaj przyrównanie sekwencji z poprzedniego zadania za pomocą programu needle. Następnie wykonaj kolejne przyrównanie tym razem dodając do polecenia przełącznik `-endweight`.

1. Na czym polega różnica między dwoma otrzymanymi przyrównaniami?
2. Które z przyrównań nazwał(a)byś semi-globalnym?