

wykład 5

Dopasowanie par sekwencji (2)

dr Jacek Śmietański

jacek.smietanski@ii.uj.edu.pl

<http://jaceksmietanski.net>

1. Macierze substytucji
2. Znaczenie kary za otwarcie przerwy - model afiniczny
3. Algorytm o liniowej złożoności pamięciowej
4. Dopasowanie lokalne - algorytm Smitha-Watermana

Macierze substytucji

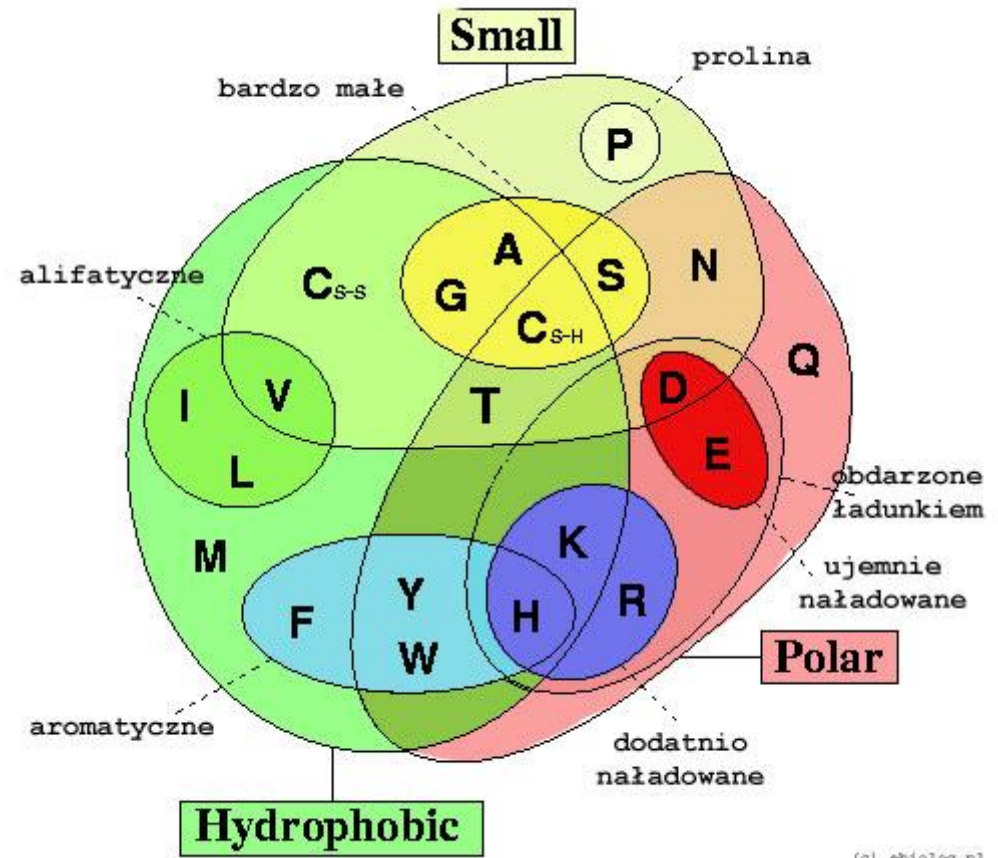
Cel konstrukcji macierzy substytucji

Uwzględnienie w dopasowaniach rzeczywistych częstości substytucji pomiędzy różniącymi się aminokwasami.

Każdy aminokwas ma inne właściwości.

Niektóre pod pewnymi względami są do siebie podobne.

Substytucje między aminokwasami podobnymi są bardziej prawdopodobne.



PAM = *Point Accepted Mutation*

Macierze oparte na modelu ewolucyjnym akceptowanych (możliwych do zaistnienia) mutacji punktowych.

1 jednostka PAM – stopień zróżnicowania ewolucyjnego, w którym zmienił się 1% aminokwasów.

Dla zdefiniowania wartości PAM, wprowadzono pojęcia:

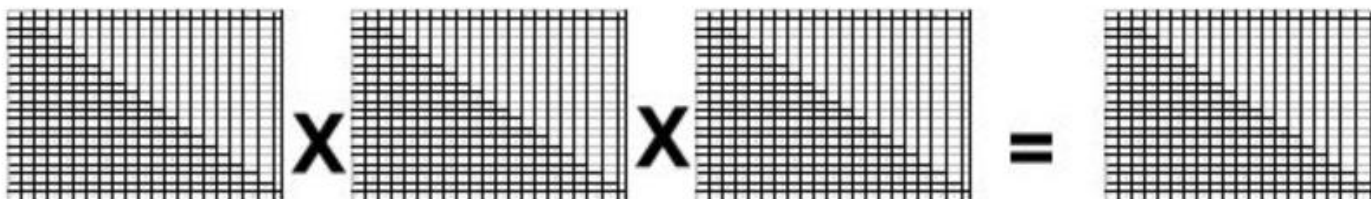
- Częstość tła – częstość zmian „przypadkowych”
- Częstość docelowa – częstość substytucji; obrazuje zmiany pojawiające się w białkach spokrewnionych

Wartości w macierzy są proporcjonalne do logarytmu ilorazu:

$$\text{częstość docelowa} / \text{częstość tła}$$

Podstawowa macierz (1PAM) zbudowana została na podstawie analizy par blisko spokrewnionych.

Macierz można ekstrapolować do większych odległości ewolucyjnych PAM.



Multiply Matrices N times to make PAM “X”; then take the Log

„Duże” PAM (np. PAM 250) stosuje się do porównywania sekwencji o dużym stopniu dywergencji ewolucyjnej.

„Małe” PAM do badania sekwencji podobnych.

Oryginalne macierze Dayhoff

W 1978 roku Dayhoff, Schwartz i Orcutt przeanalizowali 1572 możliwe mutacje pomiędzy 71 grupami blisko spokrewnionych sekwencji (>85% identyczności).

Zobacz też: <https://www.inf.ethz.ch/personal/gonnet/DarwinManual/node151.html>

		Ala	Arg	Asn	Asp	Cys	Gln	Glu	Gly	His	Ile	Leu	Lys	Met	Phe	Pro	Ser	Thr	Trp	Tyr	Val
		A	R	N	D	C	Q	E	G	H	I	L	K	M	F	P	S	T	W	Y	V
Ala	A	9867	2	9	10	3	8	17	21	2	6	4	2	6	2	22	35	32	0	2	18
Arg	R	1	9913	1	0	1	10	0	0	10	3	1	19	4	1	4	6	1	8	0	1
Asn	N	4	1	9822	36	0	4	6	6	21	3	1	13	0	1	2	20	9	1	4	1
Asp	D	6	0	42	9859	0	6	53	6	4	1	0	3	0	0	1	5	3	0	0	1
Cys	C	1	1	0	0	9973	0	0	0	1	1	0	0	0	0	1	5	1	0	3	2
Gln	Q	3	9	4	5	0	9876	27	1	23	1	3	6	4	0	6	2	2	0	0	1
Glu	E	10	0	7	56	0	35	9865	4	2	3	1	4	1	0	3	4	2	0	1	2
Gly	G	21	1	12	11	1	3	7	9935	1	0	1	2	1	1	3	21	3	0	0	5
His	H	1	8	18	3	1	20	1	0	9912	0	1	1	0	2	3	1	1	1	4	1
Ile	I	2	2	3	1	2	1	2	0	0	9872	9	2	12	7	0	1	7	0	1	33
Leu	L	3	1	3	0	0	6	1	1	4	22	9947	2	45	13	3	1	3	4	2	15
Lys	K	2	37	25	6	0	12	7	2	2	4	1	9926	20	0	3	8	11	0	1	1
Met	M	1	1	0	0	0	2	0	0	0	5	8	4	9874	1	0	1	2	0	0	4
Phe	F	1	1	1	0	0	0	0	1	2	8	6	0	4	9946	0	2	1	3	28	0
Pro	P	13	5	2	1	1	8	3	2	5	1	2	2	1	1	9926	12	4	0	0	2
Ser	S	28	11	34	7	11	4	6	16	2	2	1	7	4	3	17	9840	38	5	2	2
Thr	T	22	2	13	4	1	3	2	2	1	11	2	8	6	1	5	32	9871	0	2	9
Trp	W	0	2	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	9976	1	0
Tyr	Y	1	0	3	0	3	0	1	0	4	1	1	0	0	21	0	1	1	2	9945	1
Val	V	13	2	1	1	3	2	2	3	3	57	11	1	17	1	3	2	10	0	2	9901

Macierz prawdopodobieństwa mutacji będąca podstawą konstrukcji PAM1.

Diagonal: Min: 6.0 Max: 13.0 Mean: 8.20 Sigma: 1.81
 Rest: Min: -17.0 Max: 2.0 Mean: -6.47 Sigma: 4.14
 Matrix: Min: -17.0 Max: 13.0 Mean: -5.08 Sigma: 5.87

A	R	N	D	C	Q	E	G	H	I	L	K	M	F	P	S	T	W	Y	V	B	Z	X	..
6	-7	-4	-3	-6	-4	-2	-2	-7	-5	-6	-7	-5	-8	-2	0	-1	-13	-8	-2	-3	-3	-3	A
	8	-6	-10	-8	-2	-9	-9	-2	-5	-8	0	-4	-9	-4	-3	-6	-2	-10	-8	-7	-4	-6	R
		8	2	-11	-3	-2	-3	0	-5	-7	-1	-9	-9	-6	0	-2	-8	-4	-8	6	-3	-3	N
			8	-14	-2	2	-3	-4	-7	-12	-4	-11	-15	-8	-4	-5	-15	-11	-8	6	1	-5	D
				10	-14	-14	-9	-7	-6	-15	-14	-13	-13	-8	-3	-8	-15	-4	-6	-12	-14	-9	C
					8	1	-7	1	-8	-5	-3	-4	-13	-3	-5	-5	-13	-12	-7	-3	6	-5	Q
						8	-4	-5	-5	-9	-4	-7	-14	-5	-4	-6	-17	-8	-6	1	6	-5	E
							6	-9	-11	-10	-7	-8	-9	-6	-2	-6	-15	-14	-5	-3	-5	-5	G
								9	-9	-6	-6	-10	-6	-4	-6	-7	-7	-3	-6	-1	-1	-5	H
									8	-1	-6	-1	-2	-8	-7	-2	-14	-6	2	-6	-6	-5	I
										7	-8	1	-3	-7	-8	-7	-6	-7	-2	-9	-7	-6	L
											7	-2	-14	-6	-4	-3	-12	-9	-9	-2	-4	-5	K
												11	-4	-8	-5	-4	-13	-11	-1	-10	-5	-5	M
													9	-10	-6	-9	-4	2	-8	-10	-13	-8	F
														8	-2	-4	-14	-13	-6	-7	-4	-5	P
															6	0	-5	-7	-6	-1	-5	-3	S
																7	-13	-6	-3	-3	-6	-4	T
																	13	-5	-15	-10	-14	-11	W
																		10	-7	-6	-9	-7	Y
																			7	-8	-6	-5	V
																				6	0	-5	B
																					6	-5	Z
																						-5	X

Diagonal: Min: 3.0 Max: 18.0 Mean: 7.30 Sigma: 3.72
 Rest: Min: -10.0 Max: 7.0 Mean: -2.38 Sigma: 3.06
 Matrix: Min: -10.0 Max: 18.0 Mean: -1.46 Sigma: 4.22

A	R	N	D	C	Q	E	G	H	I	L	K	M	F	P	S	T	W	Y	V	B	Z	X	..
3	-3	0	0	-3	-1	0	1	-2	-1	-3	-2	-2	-5	1	1	2	-8	-5	0	0	0	-1	A
	8	-1	-3	-5	1	-2	-4	2	-3	-4	4	-1	-6	-1	-1	-2	2	-6	-4	-2	0	-2	R
		4	3	-5	0	2	0	2	-3	-4	1	-3	-5	-1	1	0	-5	-2	-3	3	1	-1	N
			5	-7	2	4	0	0	-3	-6	0	-4	-8	-2	0	-1	-9	-6	-3	4	3	-1	D
				13	-7	-7	-5	-4	-3	-8	-7	-7	-6	-4	0	-3	-10	0	-3	-6	-7	-4	C
					6	3	-2	4	-3	-2	0	-1	-6	0	-1	-2	-6	-6	-3	1	5	-1	Q
						5	0	0	-3	-5	-1	-3	-7	-1	-1	-1	-9	-6	-3	3	5	-1	E
							6	-3	-4	-6	-3	-4	-6	-1	1	-1	-9	-7	-2	0	-1	-2	G
								8	-4	-3	-1	-3	-3	-1	-2	-2	-4	0	-3	1	2	-1	H
									6	2	-3	2	1	-3	-2	0	-7	-2	5	-3	-3	-1	I
										7	-4	4	1	-4	-4	-3	-3	-2	2	-5	-3	-2	L
											6	1	-7	-2	-1	0	-5	-6	-4	0	0	-1	K
												9	0	-3	-2	-1	-6	-4	2	-3	-2	-1	M
													10	-6	-4	-4	0	7	-2	-6	-7	-3	F
														8	1	0	-7	-7	-2	-2	-1	-1	P
															3	2	-3	-4	-2	1	-1	0	S
																4	-7	-4	0	0	-1	-1	T
																	18	-1	-8	-7	-8	-6	W
																		11	-4	-4	-6	-3	Y
																			6	-3	-3	-1	V
																				4	3	-1	B
																					5	-1	Z
																						-1	X

Diagonal: Min: 2.0 Max: 17.0 Mean: 5.90 Sigma: 3.78
 Rest: Min: -8.0 Max: 7.0 Mean: -1.51 Sigma: 2.40
 Matrix: Min: -8.0 Max: 17.0 Mean: -0.80 Sigma: 3.35

A	R	N	D	C	Q	E	G	H	I	L	K	M	F	P	S	T	W	Y	V	B	Z	X	..
2	-2	0	0	-2	0	0	1	-1	-1	-2	-1	-1	-3	1	1	1	-6	-3	0	0	0	0	A
	6	0	-1	-4	1	-1	-3	2	-2	-3	3	0	-4	0	0	-1	2	-4	-2	-1	0	-1	R
		2	2	-4	1	1	0	2	-2	-3	1	-2	-3	0	1	0	-4	-2	-2	2	1	0	N
			4	-5	2	3	1	1	-2	-4	0	-3	-6	-1	0	0	-7	-4	-2	3	3	-1	D
				12	-5	-5	-3	-3	-2	-6	-5	-5	-4	-3	0	-2	-8	0	-2	-4	-5	-3	C
					4	2	-1	3	-2	-2	1	-1	-5	0	-1	-1	-5	-4	-2	1	3	-1	Q
						4	0	1	-2	-3	0	-2	-5	-1	0	0	-7	-4	-2	3	3	-1	E
							5	-2	-3	-4	-2	-3	-5	0	1	0	-7	-5	-1	0	0	-1	G
								6	-2	-2	0	-2	-2	0	-1	-1	-3	0	-2	1	2	-1	H
									5	2	-2	2	1	-2	-1	0	-5	-1	4	-2	-2	-1	I
										6	-3	4	2	-3	-3	-2	-2	-1	2	-3	-3	-1	L
											5	0	-5	-1	0	0	-3	-4	-2	1	0	-1	K
												6	0	-2	-2	-1	-4	-2	2	-2	-2	-1	M
													9	-5	-3	-3	0	7	-1	-4	-5	-2	F
														6	1	0	-6	-5	-1	-1	0	-1	P
															2	1	-2	-3	-1	0	0	0	S
																3	-5	-3	0	0	-1	0	T
																	17	0	-6	-5	-6	-4	W
																		10	-2	-3	-4	-2	Y
																			4	-2	-2	-1	V
																				3	2	-1	B
																					3	-1	Z
																						-1	X

Diagonal: Min: 1.0 Max: 22.0 Mean: 6.60 Sigma: 5.17
 Rest: Min: -9.0 Max: 9.0 Mean: -1.57 Sigma: 2.77
 Matrix: Min: -9.0 Max: 22.0 Mean: -0.80 Sigma: 3.88

A	R	N	D	C	Q	E	G	H	I	L	K	M	F	P	S	T	W	Y	V	B	Z	X	..
2	-1	0	0	-2	0	0	2	-1	0	-2	-1	-1	-4	1	1	1	-6	-4	0	0	0	0	A
	7	0	-1	-4	2	-1	-2	2	-2	-3	4	0	-5	0	0	-1	3	-5	-3	0	0	-1	R
		2	2	-4	1	2	1	2	-2	-3	1	-2	-4	0	1	0	-5	-2	-2	2	1	0	N
			4	-6	2	4	1	1	-2	-4	0	-3	-6	-1	0	0	-7	-5	-2	3	3	-1	D
				15	-6	-6	-4	-4	-3	-7	-6	-6	-5	-3	0	-2	-9	1	-2	-5	-6	-3	C
					4	3	-1	3	-2	-2	1	-1	-5	0	0	-1	-5	-4	-2	2	3	0	Q
						4	0	1	-2	-4	0	-2	-6	0	0	0	-8	-5	-2	3	3	-1	E
							5	-2	-3	-4	-2	-3	-5	0	1	0	-8	-6	-1	1	0	-1	G
								7	-2	-2	0	-2	-2	0	-1	-1	-3	0	-2	1	2	0	H
									5	3	-2	3	1	-2	-1	0	-6	-1	4	-2	-2	-1	I
										7	-3	4	3	-3	-3	-2	-2	0	2	-4	-3	-1	L
											5	0	-6	-1	0	0	-4	-5	-2	1	1	-1	K
												6	1	-2	-2	-1	-5	-2	2	-2	-2	-1	M
													11	-5	-4	-3	1	9	-1	-5	-5	-2	F
														6	1	1	-6	-5	-1	0	0	-1	P
															1	1	-3	-3	-1	1	0	0	S
																2	-6	-3	0	0	0	0	T
																	22	0	-7	-6	-6	-4	W
																		12	-3	-4	-5	-2	Y
																			5	-2	-2	0	V
																				3	2	0	B
																					3	-1	Z
																						-1	X

Powstały w oparciu o bazę BLOCKS – dopasowanie sekwencji daleko spokrewnionych
(oszacowanie częstotliwości docelowych, bez modelu ewolucyjnego)

Rodzina macierzy:
różnice indeksu związane są z maksymalnym stopniem identyczności sekwencji wziętych do obliczeń.

Przykłady:

BLOSUM90 – do analizy sekwencji blisko spokrewnionych

BLOSUM62 – najczęściej używana

BLOSUM30 – do analizy odległych ewolucyjnie sekwencji

BLOSUM 45
PAM 250

BLOSUM 62
PAM 160

BLOSUM 90
PAM 100



bardziej zróżnicowane



mniej zróżnicowane

Zastosowanie:
BLOSUM – lokalne przyrównania sekwencji
PAM – konstrukcja drzew filogenetycznych

X=0

C 12

S 0 2

T -2 1 3

P -3 1 0 6

A -2 1 1 1 2

G -3 1 0 -1 1 5

N -4 1 0 -1 0 0 2

D -5 0 0 -1 0 1 2 4

E -5 0 0 -1 0 0 1 3 4

Q -5 -1 -1 0 0 -1 1 2 2 4

H -3 -1 -1 0 -1 -2 2 1 1 3 6

R -4 0 -1 0 -2 -3 0 -1 -1 1 2 6

K -5 0 0 -1 -1 -2 1 0 0 1 0 3 5

M -5 -2 -1 -2 -1 -3 -2 -3 -2 -1 -2 0 0 6

I -2 -1 0 -2 -1 -3 -2 -2 -2 -2 -2 2 5

L -6 -3 -2 -3 -2 -4 -3 -4 -3 -2 -2 -3 4 2 6

V -2 -1 0 -1 0 -1 -2 -2 -2 -2 -2 2 4 2 4

F -4 -3 -3 -5 -4 -5 -4 -6 -5 -5 -2 -4 -5 0 1 2 -1 9

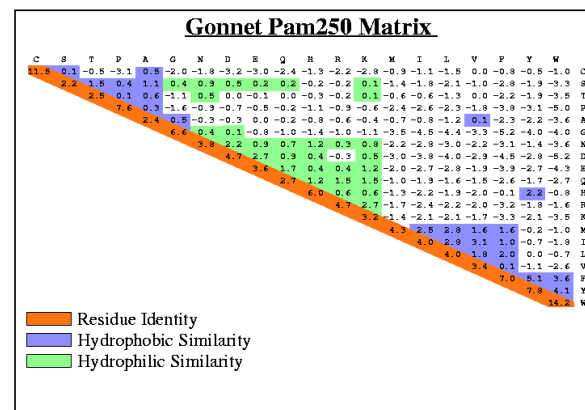
W 0 -3 -3 -5 -3 -5 -2 -4 -4 -4 0 -4 -4 -2 -1 -1 -2 7 10

Y -8 -2 -5 -6 -6 -7 -4 -7 -7 -5 -3 2 -3 -4 -5 -2 -6 0 0 17

C S T P A G N D E Q H R K M I L V F W Y

PAM 250 matrix

(c) David Gilbert, 2003 [Sequence Comparison]



33

BLOSUM 45

Diagonal: Min: 4.0 Max: 15.0 Mean: 7.05 Sigma: 2.79
 Rest: Min: -5.0 Max: 3.0 Mean: -1.34 Sigma: 1.60
 Matrix: Min: -5.0 Max: 15.0 Mean: -0.54 Sigma: 3.02

A	R	N	D	C	Q	E	G	H	I	L	K	M	F	P	S	T	W	Y	V	B	Z	X	..
5	-2	-1	-2	-1	-1	-1	0	-2	-1	-1	-1	-1	-2	-1	1	0	-2	-2	0	-1	-1	0	A
	7	0	-1	-3	1	0	-2	0	-3	-2	3	-1	-2	-2	-1	-1	-2	-1	-2	-1	0	-1	R
		6	2	-2	0	0	0	1	-2	-3	0	-2	-2	-2	1	0	-4	-2	-3	4	0	-1	N
			7	-3	0	2	-1	0	-4	-3	0	-3	-4	-1	0	-1	-4	-2	-3	5	1	-1	D
				12	-3	-3	-3	-3	-3	-2	-3	-2	-2	-4	-1	-1	-5	-3	-1	-2	-3	-2	C
					6	2	-2	1	-2	-2	1	0	-4	-1	0	-1	-2	-1	-3	0	4	-1	Q
						6	-2	0	-3	-2	1	-2	-3	0	0	-1	-3	-2	-3	1	4	-1	E
							7	-2	-4	-3	-2	-2	-3	-2	0	-2	-2	-3	-3	-1	-2	-1	G
								10	-3	-2	-1	0	-2	-2	-1	-2	-3	2	-3	0	0	-1	H
									5	2	-3	2	0	-2	-2	-1	-2	0	3	-3	-3	-1	I
										5	-3	2	1	-3	-3	-1	-2	0	1	-3	-2	-1	L
											5	-1	-3	-1	-1	-1	-2	-1	-2	0	1	-1	K
												6	0	-2	-2	-1	-2	0	1	-2	-1	-1	M
													8	-3	-2	-1	1	3	0	-3	-3	-1	F
														9	-1	-1	-3	-3	-3	-2	-1	-1	P
															4	2	-4	-2	-1	0	0	0	S
																5	-3	-1	0	0	-1	0	T
																	15	3	-3	-4	-2	-2	W
																		8	-1	-2	-2	-1	Y
																			5	-3	-3	-1	V
																				4	2	-1	B
																					4	-1	Z
																						-1	X

BLOSUM 62

Diagonal: Min: 4.0 Max: 11.0 Mean: 5.80 Sigma: 1.90
 Rest: Min: -4.0 Max: 3.0 Mean: -1.43 Sigma: 1.51
 Matrix: Min: -4.0 Max: 11.0 Mean: -0.74 Sigma: 2.63

A	R	N	D	C	Q	E	G	H	I	L	K	M	F	P	S	T	W	Y	V	B	Z	X	..
4	-1	-2	-2	0	-1	-1	0	-2	-1	-1	-1	-1	-2	-1	1	0	-3	-2	0	-2	-1	0	A
	5	0	-2	-3	1	0	-2	0	-3	-2	2	-1	-3	-2	-1	-1	-3	-2	-3	-1	0	-1	R
		6	1	-3	0	0	0	1	-3	-3	0	-2	-3	-2	1	0	-4	-2	-3	3	0	-1	N
			6	-3	0	2	-1	-1	-3	-4	-1	-3	-3	-1	0	-1	-4	-3	-3	4	1	-1	D
				9	-3	-4	-3	-3	-1	-1	-3	-1	-2	-3	-1	-1	-2	-2	-1	-3	-3	-2	C
					5	2	-2	0	-3	-2	1	0	-3	-1	0	-1	-2	-1	-2	0	3	-1	Q
						5	-2	0	-3	-3	1	-2	-3	-1	0	-1	-3	-2	-2	1	4	-1	E
							6	-2	-4	-4	-2	-3	-3	-2	0	-2	-2	-3	-3	-1	-2	-1	G
								8	-3	-3	-1	-2	-1	-2	-1	-2	-2	2	-3	0	0	-1	H
									4	2	-3	1	0	-3	-2	-1	-3	-1	3	-3	-3	-1	I
										4	-2	2	0	-3	-2	-1	-2	-1	1	-4	-3	-1	L
											5	-1	-3	-1	0	-1	-3	-2	-2	0	1	-1	K
												5	0	-2	-1	-1	-1	-1	1	-3	-1	-1	M
													6	-4	-2	-2	1	3	-1	-3	-3	-1	F
														7	-1	-1	-4	-3	-2	-2	-1	-2	P
															4	1	-3	-2	-2	0	0	0	S
																5	-2	-2	0	-1	-1	0	T
																	11	2	-3	-4	-3	-2	W
																		7	-1	-3	-2	-1	Y
																			4	-3	-2	-1	V
																				4	1	-1	B
																					4	-1	Z
																						-1	X

BLOSUM 90

Diagonal: Min: 5.0 Max: 11.0 Mean: 6.70 Sigma: 1.60
 Rest: Min: -6.0 Max: 3.0 Mean: -2.32 Sigma: 1.84
 Matrix: Min: -6.0 Max: 11.0 Mean: -1.46 Sigma: 3.22

A	R	N	D	C	Q	E	G	H	I	L	K	M	F	P	S	T	W	Y	V	B	Z	X	..
5	-2	-2	-3	-1	-1	-1	0	-2	-2	-2	-1	-2	-3	-1	1	0	-4	-3	-1	-2	-1	-1	A
	6	-1	-3	-5	1	-1	-3	0	-4	-3	2	-2	-4	-3	-1	-2	-4	-3	-3	-2	0	-2	R
		7	1	-4	0	-1	-1	0	-4	-4	0	-3	-4	-3	0	0	-5	-3	-4	4	-1	-2	N
			7	-5	-1	1	-2	-2	-5	-5	-1	-4	-5	-3	-1	-2	-6	-4	-5	4	0	-2	D
				9	-4	-6	-4	-5	-2	-2	-4	-2	-3	-4	-2	-2	-4	-4	-2	-4	-5	-3	C
					7	2	-3	1	-4	-3	1	0	-4	-2	-1	-1	-3	-3	-3	-1	4	-1	Q
						6	-3	-1	-4	-4	0	-3	-5	-2	-1	-1	-5	-4	-3	0	4	-2	E
							6	-3	-5	-5	-2	-4	-5	-3	-1	-3	-4	-5	-5	-2	-3	-2	G
								8	-4	-4	-1	-3	-2	-3	-2	-2	-3	1	-4	-1	0	-2	H
									5	1	-4	1	-1	-4	-3	-1	-4	-2	3	-5	-4	-2	I
										5	-3	2	0	-4	-3	-2	-3	-2	0	-5	-4	-2	L
											6	-2	-4	-2	-1	-1	-5	-3	-3	-1	1	-1	K
												7	-1	-3	-2	-1	-2	-2	0	-4	-2	-1	M
													7	-4	-3	-3	0	3	-2	-4	-4	-2	F
														8	-2	-2	-5	-4	-3	-3	-2	-2	P
															5	1	-4	-3	-2	0	-1	-1	S
																6	-4	-2	-1	-1	-1	-1	T
																	11	2	-3	-6	-4	-3	W
																		8	-3	-4	-3	-2	Y
																			5	-4	-3	-2	V
																				4	0	-2	B
																					4	-1	Z
																						-2	X

BLOSUM 62 - wizualizacja

	C	S	T	P	A	G	N	D	E	Q	H	R	K	M	I	L	V	F	Y	W	
C	9																				C
S	-1	4																			S
T	-1	1	5																		T
P	-3	-1	-1	7																	P
A	0	1	0	-1	4																A
G	-3	0	-2	-2	0	6															G
N	-3	1	0	-2	-2	0	6														N
D	-3	0	-1	-1	-2	-1	1	6													D
E	-4	0	-1	-1	-1	-2	0	2	5												E
Q	-3	0	-1	-1	-1	-2	0	0	2	5											Q
H	-3	-1	-2	-2	-2	-2	1	-1	0	0	8										H
R	-3	-1	-1	-2	-1	-2	0	-2	0	1	0	5									R
K	-3	0	-1	-1	-1	-2	0	-1	1	1	-1	2	5								K
M	-1	-1	-1	-2	-1	-3	-2	-3	-2	0	-2	-1	-1	5							M
I	-1	-2	-1	-3	-1	-4	-3	-3	-3	-3	-3	-3	-3	1	4						I
L	-1	-2	-1	-3	-1	-4	-3	-4	-3	-2	-3	-2	-2	2	2	4					L
V	-1	-2	0	-2	0	-3	-3	-3	-2	-2	-3	-3	-2	1	3	1	4				V
F	-2	-2	-2	-4	-2	-3	-3	-3	-3	-3	-1	-3	-3	0	0	0	-1	6			F
Y	-2	-2	-2	-3	-2	-3	-2	-3	-2	-1	2	-2	-2	-1	-1	-1	-1	3	7		Y
W	-2	-3	-2	-4	-3	-2	-4	-4	-3	-2	-2	-3	-3	-1	-3	-2	-3	1	2	11	W

```
from Bio.SubsMat import MatrixInfo
```

```
available_matrices = ['benner6', 'benner22', 'benner74', 'blos...  
benner6 = {('A', 'A'): 2.5, ('A', 'C'): -1.7, ('A', 'P'): 1.1,...  
benner22 = {('A', 'A'): 2.5, ('A', 'C'): -1.2, ('A', 'P'): 0.8...  
benner74 = {('A', 'A'): 2.4, ('A', 'C'): 0.3, ('A', 'P'): 0.4,...  
blosum100 = {('A', 'A'): 5, ('B', 'A'): -3, ('B', 'B'): 4, ('B...  
blosum30 = {('A', 'A'): 4, ('B', 'A'): 0, ('B', 'B'): 5, ('B',...  
blosum35 = {('A', 'A'): 5, ('B', 'A'): -1, ('B', 'B'): 5, ('B'...  
blosum40 = {('A', 'A'): 5, ('B', 'A'): -1, ('B', 'B'): 5, ('B'...  
blosum45 = {('A', 'A'): 5, ('B', 'A'): -1, ('B', 'B'): 4, ('B'...  
blosum50 = {('A', 'A'): 5, ('B', 'A'): -2, ('B', 'B'): 5, ('B'...  
blosum55 = {('A', 'A'): 5, ('B', 'A'): -2, ('B', 'B'): 5, ('B'...  
blosum60 = {('A', 'A'): 4, ('B', 'A'): -2, ('B', 'B'): 4, ('B'...  
blosum62 = {('A', 'A'): 4, ('B', 'A'): -2, ('B', 'B'): 4, ('B'...  
blosum65 = {('A', 'A'): 4, ('B', 'A'): -2, ('B', 'B'): 4, ('B'...  
blosum70 = {('A', 'A'): 4, ('B', 'A'): -2, ('B', 'B'): 4, ('B'...  
blosum75 = {('A', 'A'): 4, ('B', 'A'): -2, ('B', 'B'): 4, ('B'...  
blosum80 = {('A', 'A'): 5, ('B', 'A'): -2, ('B', 'B'): 4, ('B'...  
blosum85 = {('A', 'A'): 5, ('B', 'A'): -2, ('B', 'B'): 4, ('B'...  
blosum90 = {('A', 'A'): 5, ('B', 'A'): -2, ('B', 'B'): 4, ('B'...  
blosum95 = {('A', 'A'): 5, ('B', 'A'): -3, ('B', 'B'): 4, ('B'...  
feng = {('A', 'A'): 6, ('A', 'C'): 2, ('A', 'P'): 5, ('A', 'S'...
```

Najczęściej używane macierze:

- **BLOSUM62** (standard dla programu BLAST)
- **PAM250**

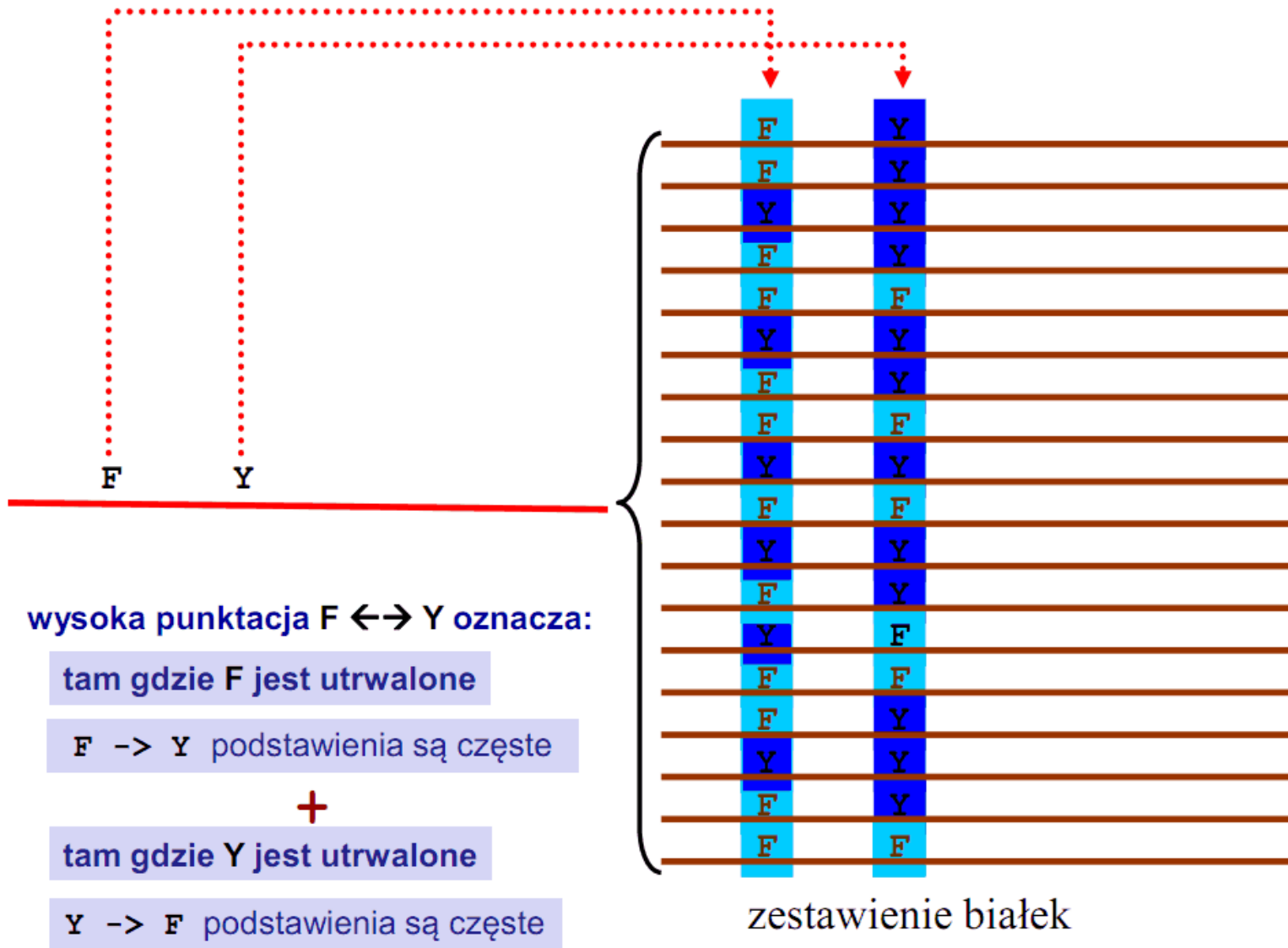
Interpretacja: częste podstawienia (1)

Dayhoff PAM 250 Matrix

	A	R	N	D	C	Q	E	G	H	I	L	K	M	F	P	Y	V	B	Z	X
A	2	-2	0	0	-2	0	0	1	-1	-1	-2	-1	-1	-4	1	-3	0	0	0	0
R	-2	6	0	-1	-4	1	-1	-3	2	-2	-3	3	0	-4	0	-4	-2	-1	0	0
N	0	0	2	2	-4	1	1	0	2	-2	-3	1	-2	-4	-1	-2	-2	2	1	0
D	0	-1	2	4	-5	2	3	1	1	-2	-4	0	-3	-6	-1	-4	-2	3	3	0
C	-2	-4	-4	-5	12	-5	-5	-3	-3	-2	-6	-5	-5	-4	-3	0	-2	-4	-5	0
Q	0	1	1	2	-5	4	2	-1	3	-2	-2	1	-1	-5	0	-1	-1	-5	-4	0
E	0	-1	1	3	-5	2	4	0	1	-2	-3	0	-2	-5	-1	0	0	-7	-4	0
G	1	-3	0	1	-3	-1	0	5	-2	-3	-4	-2	-3	-5	-1	1	0	-7	-5	0
H	-1	2	2	1	-3	3	1	-2	6	-2	-2	0	-2	-2	0	-1	-1	-3	0	0
I	-1	-2	-2	-2	-2	-2	-2	-3	-2	5	2	-2	2	1	-2	-1	0	-5	-1	0
L	-2	-3	-3	-4	-6	-2	-3	-4	-2	2	6	-3	4	2	-3	-3	-2	-2	-1	0
K	-1	3	1	0	-5	1	0	-2	0	-2	4	5	0	-5	-1	0	0	-3	-4	0
M	-1	0	2	3	-5	-1	-2	-3	-2	2	2	0	6	0	-2	-2	-1	-4	2	0
F	-1	0	2	3	-5	-1	-2	-3	-2	2	2	0	6	0	-2	-2	-1	-4	2	0
P	-1	0	2	3	-5	-1	-2	-3	-2	2	2	0	6	0	-2	-2	-1	-4	2	0
S	0	0	-1	1	1	-1	-1	-1	-3	0	-2	-3	-1	-2	-5	6	1	0	-6	0
T	0	0	-1	1	1	-1	-1	-1	-3	0	-2	-3	-1	-2	-5	6	1	0	-6	0
W	-6	2	-4	-7	-8	-5	-7	-7	-3	-5	-2	-3	-4	0	-6	-2	-5	17	0	0
Y	-3	-4	-2	-4	0	-4	-4	-5	0	-1	-1	-4	-2	7	-5	-3	-3	0	10	0
V	0	-2	-2	-2	-2	-2	-2	-1	-2	4	2	-2	2	-1	-1	-1	0	-6	-2	0
B	0	-1	2	3	-4	1	2	0	1	-2	-3	1	-2	-5	-1	0	0	-5	-3	0
Z	0	0	1	3	-5	3	3	-1	2	-2	-3	0	-2	-5	0	0	-1	-6	-4	0
X	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Macierz PAM 250, podstawienie fenyloalanina (F) – tyrozyna (Y)

Interpretacja: częste podstawienia (2)



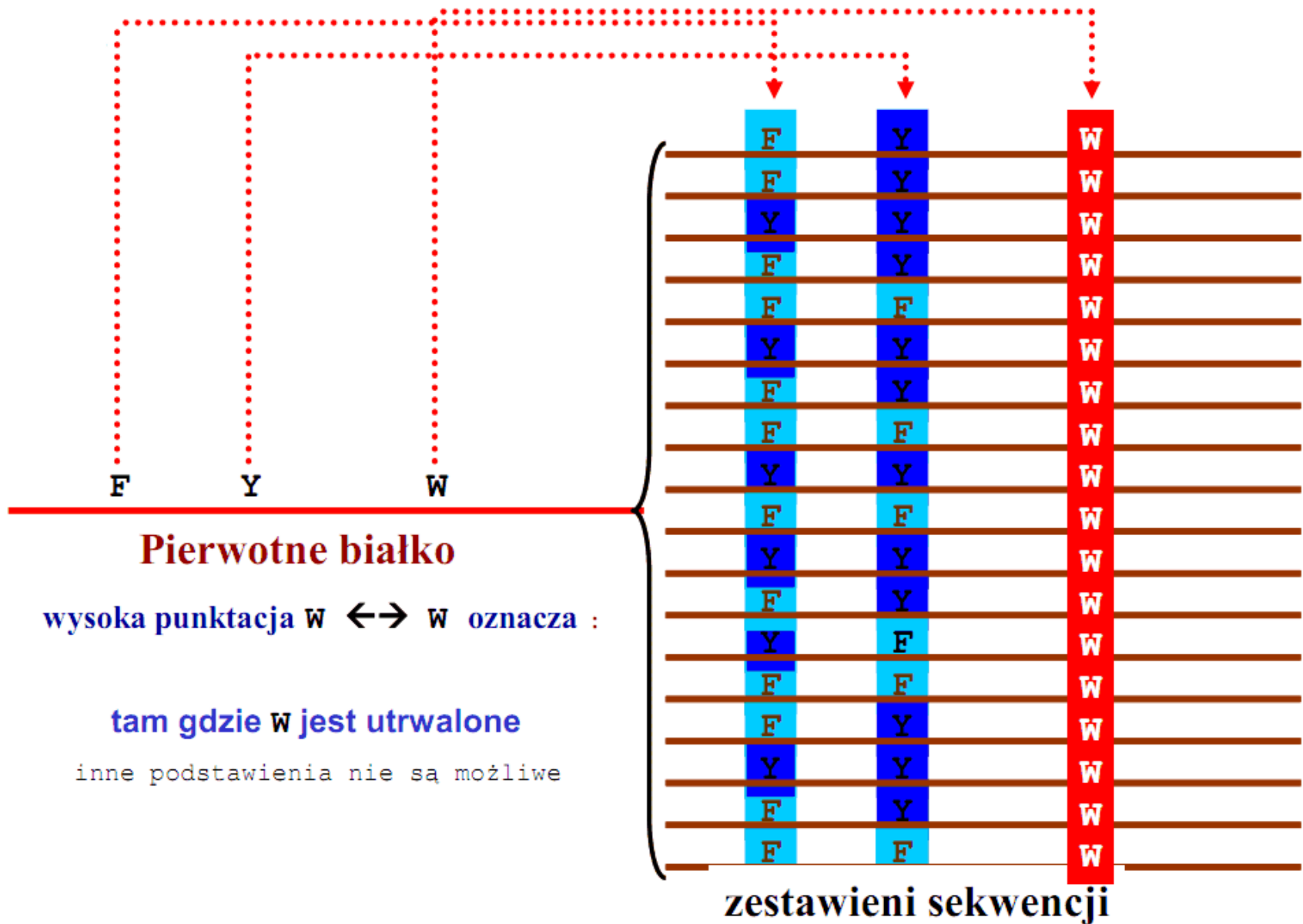
Interpretacja: utrwalona pozycja (1)

Dayhoff PAM 250 Matrix

	A	R	N	D	C	Q	E	G	H	I	L	K	M	F	P	S	T	W	Y	V	B	Z	X
A	2	-2	0	0	-2	0	0	1	-1	-1	-2	-1	-1	-4	1	1	1	-6	-3	0	0	0	0
R	-2	6	0	-1	-4	1	-1	-3	2	-2	-3	3	0	-4	0	0	-1	2	-4	-2	-1	0	0
N	0	0	2	2	-4	1	1	0	2	-2	-3	1	-2	-4	-1	1	0	-4	-2	-2	2	1	0
D	0	-1	2	4	-5	2	3	1	1	-2	-4	0	-3	-6	-1	0	0	-7	-4	-2	3	3	0
C	-2	-4	-4	-5	12	-5	-5	-3	-3	-2	-6	-5	-5	-4	-3	0	-2	-8	0	-2	-4	-5	0
Q	0	1	1	2	-5	4	2	-1	3	-2	-2	1	-1	-5	0	-1	-1	-5	-4	-2	1	3	0
E	0	-1	1	3	-5	2	4	0	1	-2	-3	0	-2	-5	-1	0	0	-7	-4	-2	2	3	0
G	1	-3	0	1	-3	-1	0	5	-2	-3	-4	-2	-3	-5	-1	1	0	-7	-5	-1	0	-1	0
H	-1	2	2	1	-3	3	1	-2	6	-2	-2	0	-2	-2	0	-1	-1	-3	0	-2	1	2	0
I	-1	-2	-2	-2	-2	-2	-2	-3	-2	5	2	-2	2	1	-2	-1	0	-5	-1	4	-2	-2	0
L	-2	-3	-3	-4	-6	-2	-3	-4	-2	-2	6	-3	4	2	-3	-3	-2	-2	-1	2	-3	-3	0
K	-1	3	1	0	-5	1	0	-2	0	-2	4	5	0	-5	-1	0	0	-3	-4	-2	1	0	0
M	-1	0	-2	-3	-5	-1	-2	-3	-2	2	2	0	6	0	-2	-2	-1	-4	2	2	-2	-2	0
F	-4	-4	-4	-6	-4	-5	-5	-5	-2	1	2	-5	0	9	-5	-3	-3	0	7	-1	-5	-5	0
P	1	0	-1	-1	-3	0	-1	-1	0	-2	-3	-1	-2	-5	6	1	0	-6	-5	-1	-1	0	0
S	1	0	1	0	0	-1	1	1	-1	-1	-3	0	-2	-3	1	2	1	-2	-3	-1	0	0	0
T	1	-1	0	0	-2	-2	0	0	-1	0	-2	0	-1	-3	0	1	3	-5	-3	0	0	-1	0
W	-6	2	-4	-7	-8	-5	-7	-7	-3	-5	-2	-3	-4	0	-6	-2	-5	17	0	-6	-5	-6	0
Y	-3	-4	-2	-4	0	-4	-4	-5	0	-1	-1	-4	-2	7	-5	-3	-3	0	10	-2	-3	-4	0
V	0	-2	-2	-2	-2	-2	-2	-1	-2	4	2	-2	2	-1	-1	-1	0	-6	-2	4	-2	-2	0
B	0	-1	2	3	-4	1	2	0	1	-2	-3	1	-2	-5	-1	0	0	-5	-3	-2	2	2	0
Z	0	0	1	3	-5	3	3	-1	2	-2	-3	0	-2	-5	0	0	-1	-6	-4	-2	2	3	0
X	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

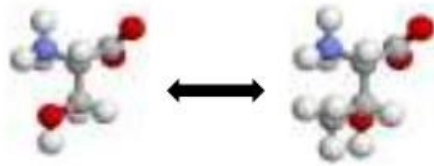
Macierz PAM 250, tryptofan (W)

Interpretacja: utrwalona pozycja (2)



Przykłady innych częstych podstawień

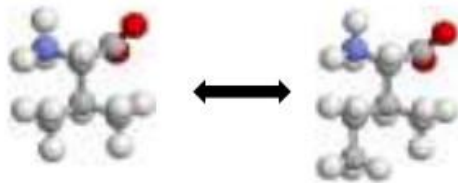
Małe, polarne



S, Ser

T, Thr

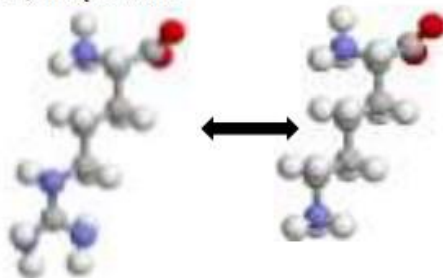
Małe, hydrofobowe



V,Val

I,Ile

Duże, naładowane



R, Arg

K, Lys

C Cys	12																					
S Ser	0	2																				
T Thr	-2	1	3																			
P Pro	-3	1	0	6																		
A Ala	-2	1	1	1	2																	
G Gly	-3	1	0	-1	1	5																
N Asn	-4	1	0	-1	0	0	2															
D Asp	-5	0	0	-1	0	1	2	4														
E Glu	-5	0	0	-1	0	0	1	3	4													
Q Gln	-5	-1	-1	0	0	-1	1	2	2	4												
H His	-3	-1	-1	0	-1	-2	2	1	1	3	6											
R Arg	-4	0	-1	0	-2	-3	0	-1	-1	1	2	6										
K Lys	-5	0	0	-1	-1	-2	1	0	0	1	0	3	5									
M Met	-5	-2	-1	-2	-1	-3	-2	-3	-2	-1	-2	0	0	6								
I Ile	-2	-1	0	-2	-1	-3	-2	-2	-2	-2	-2	-2	-2	2	5							
L Leu	-6	-3	-2	-3	-2	-4	-3	-4	-3	-2	-2	-3	-3	4	2	6						
V Val	-2	-1	0	-1	0	-1	-2	-2	-2	-2	-2	-2	-2	2	4	2	4					
F Phe	-4	-3	-3	-5	-5	-5	-4	-6	-5	-5	-2	-4	-5	0	1	2	-1	9				
Y Tyr	0	-3	-3	-5	-3	-5	-2	-4	-4	-4	0	-4	-4	-2	-1	-1	-2	7	10			
W Trp	-8	-2	-5	-6	-6	-7	-4	-7	-7	-5	-3	2	-3	-4	-5	-2	-6	0	0	17		
	C	S	T	P	A	G	N	D	E	Q	H	R	K	M	I	L	V	F	Y	W		

Źródło: <http://www.cryst.bbk.ac.uk/pps97/assignments/projects/leluk/project.htm>

Algorytm Needlemana-Wunscha **model afiniczny**

Niech:

k – długość przerwy;

G_p – całkowita kara za przerwę

Model liniowy:

$$G_p = k * G \quad (\text{każda przerwa traktowana jest tak samo})$$

Model afiniczny:

$$G_p = G_o + k * G_e^*$$

G_o (*gap opening*) – otwarcie przerwy;

G_e (*gap extension*) – wydłużenie przerwy;

* można się też spotkać z notacją: $G_p = G_o + (k-1) * G_e$
wówczas $G_o > G_e$

S1 = ATCTA

S2 = ATTTTTA

1) **AT-C-TA**
 ATTTTTA $4-1-2-2 = -1$

2) **ATC--TA**
 ATTTTTA $4-1-2-1 = 0$

3) **AT-C--TA**
 ATT-TTTA $4-2-2-2-1 = -3$

Punktacja

zgodność: +1

niezgodność: -1

przerwa: -1

+otwarcie przerwy: -1

Wg alternatywnej notacji

zgodność: +1

niezgodność: -1

otwarcie przerwy: -2

wydłużenie przerwy: -1

Tworzymy trzy macierze: H , E i F :

$$H_{i,j} = \max \begin{cases} H_{i-1,j-1} + s(a_i, b_j) \\ E_{i,j} \\ F_{i,j} \end{cases}$$

$$E_{i,j} = \max \begin{cases} H_{i-1,j} + Go + Ge \\ E_{i-1,j} + Ge \end{cases}$$

$$F_{i,j} = \max \begin{cases} H_{i,j-1} + Go + Ge \\ F_{i,j-1} + Ge \end{cases}$$

$$H_{0,0} = 0$$

$$H_{i,0} = G_o + i^*G_e$$

$$H_{0,j} = G_o + j^*G_e$$

$$E_{0,0} = -\infty$$

$$E_{i,0} = G_o + i^*G_e$$

$$E_{0,j} = -\infty$$

$$F_{0,0} = -\infty$$

$$F_{i,0} = -\infty$$

$$F_{0,j} = G_o + j^*G_e$$

Zgodnie z kierunkiem, z którego mogliśmy przyjść:

Z $H_{i,j}$ idziemy do $H_{i-1,j-1}$, $E_{i,j}$ lub $F_{i,j}$

Z $E_{i,j}$ idziemy do $E_{i-1,j}$, lub $H_{i-1,j}$

Z $F_{i,j}$ idziemy do $F_{i,j-1}$, lub $H_{i,j-1}$

$$H_{i,j} = \max \begin{cases} H_{i-1,j-1} + s(a_i, b_j) \\ E_{i,j} \\ F_{i,j} \end{cases}$$

$$E_{i,j} = \max \begin{cases} H_{i-1,j} + Go + Ge \\ E_{i-1,j} + Ge \end{cases}$$

$$F_{i,j} = \max \begin{cases} H_{i,j-1} + Go + Ge \\ F_{i,j-1} + Ge \end{cases}$$

Model afiniczny – przykład

Z $H_{i,j}$ idziemy do $H_{i-1,j-1}$, $E_{i,j}$ lub $F_{i,j}$
 Z $E_{i,j}$ idziemy do $E_{i-1,j}$ lub $H_{i-1,j}$
 Z $F_{i,j}$ idziemy do $F_{i,j-1}$ lub $H_{i,j-1}$

		V	L	S	P	A	
$H[][]$		0	-8	-9	-10	-11	-12
V		-8	↖ 2	-6	-7	-8	-9
S		-9	-6	↖ 1	-4	↖ -8	-9
A		-10	-7	-7	0	-5	↖ -6

		V	L	S	P	A	
$E[][]$		$-\infty$	$-\infty$	$-\infty$	$-\infty$	$-\infty$	
V		-8	-16	-17	-18	-19	-20
S		-9	-6	-14	-15	-16	-17
A		-10	-7	-7	-12	-16	-17

		V	L	S	P	A	
$F[][]$		$-\infty$	-8	-9	-10	-11	-12
V		$-\infty$	-16	←-6	←-7	-8	-9
S		$-\infty$	-17	-14	←-7	←-8	-9
A		$-\infty$	-18	-15	-15	-8	-9

	0	1	2	3	4
0	V	S	-	-	A
1	V	L	S	P	A

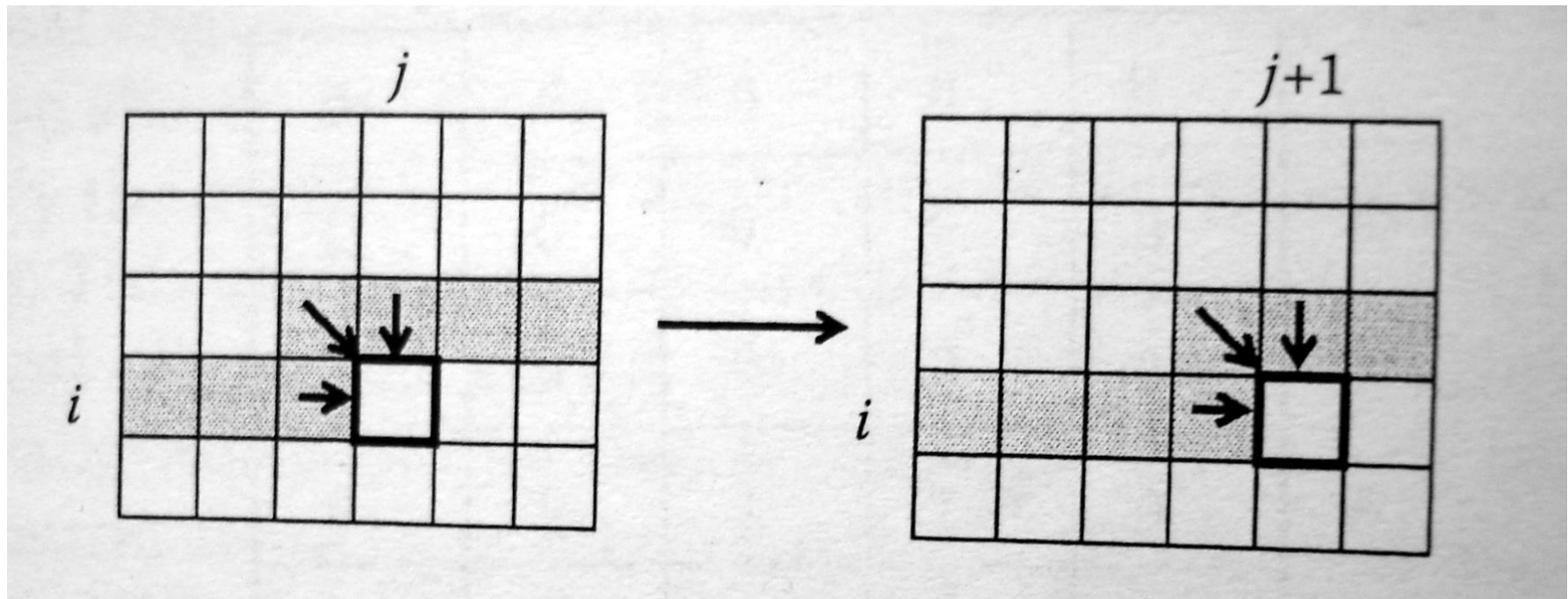
	0	1	2	3	4
0	V	-	-	S	A
1	V	L	S	P	A

Algorytm Needlemana-Wunscha

liniowa złożoność pamięciowa

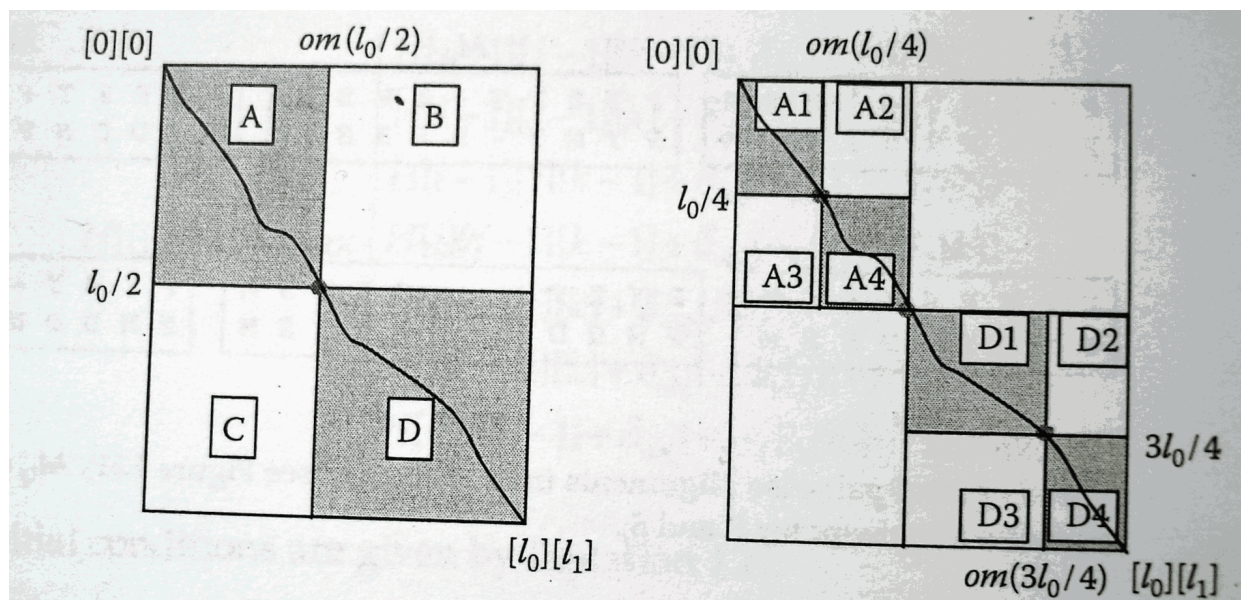
Przydatny, gdy chcemy porównywać długie sekwencje.

np. dla porównania dwóch sekwencji o długości 1000000 nukleotydów każda, potrzeba ok. Tb pamięci dla zapamiętania samej macierzy.



Pamiętamy tylko ostatni wiersz ($n+2$ komórki).

Poszukiwanie dopasowania - metoda

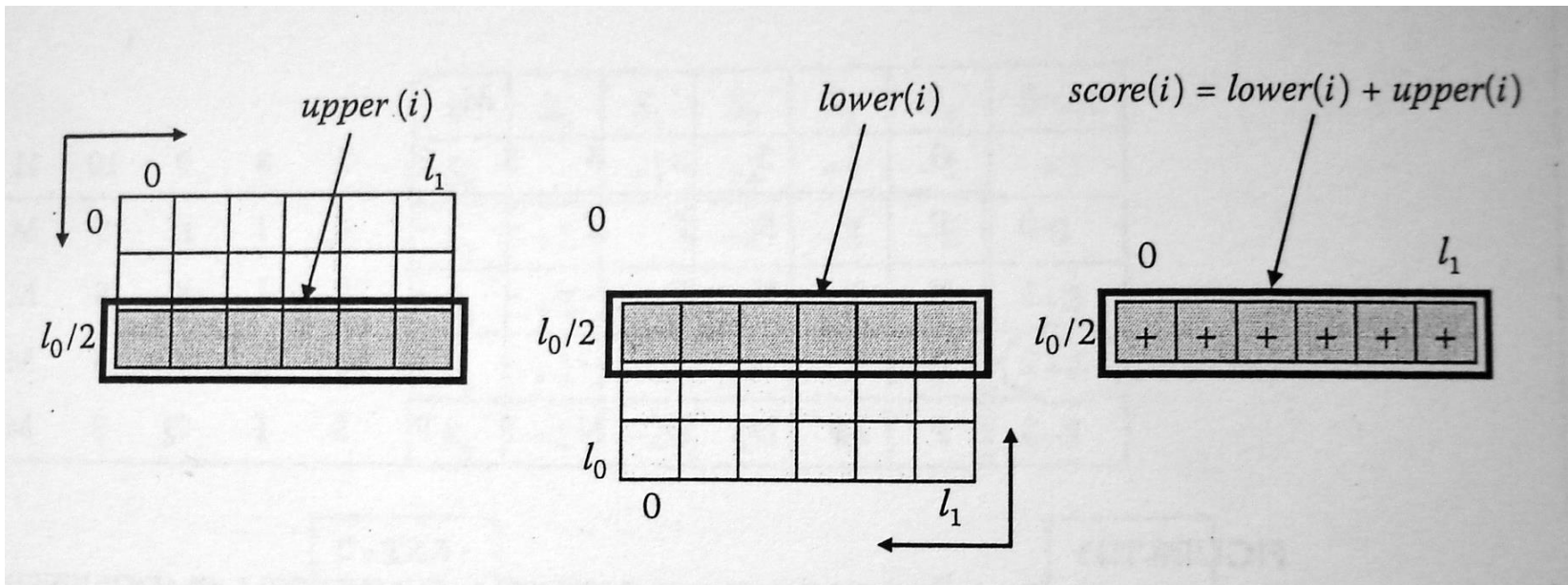


Metoda „dziel i zwyciężaj”. Stosujemy rekurencyjnie:

Idea pojedynczego etapu: W ustalonym wierszu na wysokości połowy rozważanej macierzy znajdujemy punkt o największej wartości dopasowania (szczegóły na następnym slajdzie); w tym miejscu dzielimy macierz na 4 prostokąty.

Rozważamy oddzielnie lewy górny i prawy dolny prostokąt, dla każdego z nich powtarzając powyższą procedurę.

Poszukiwanie dopasowania – wybór punktu podziału

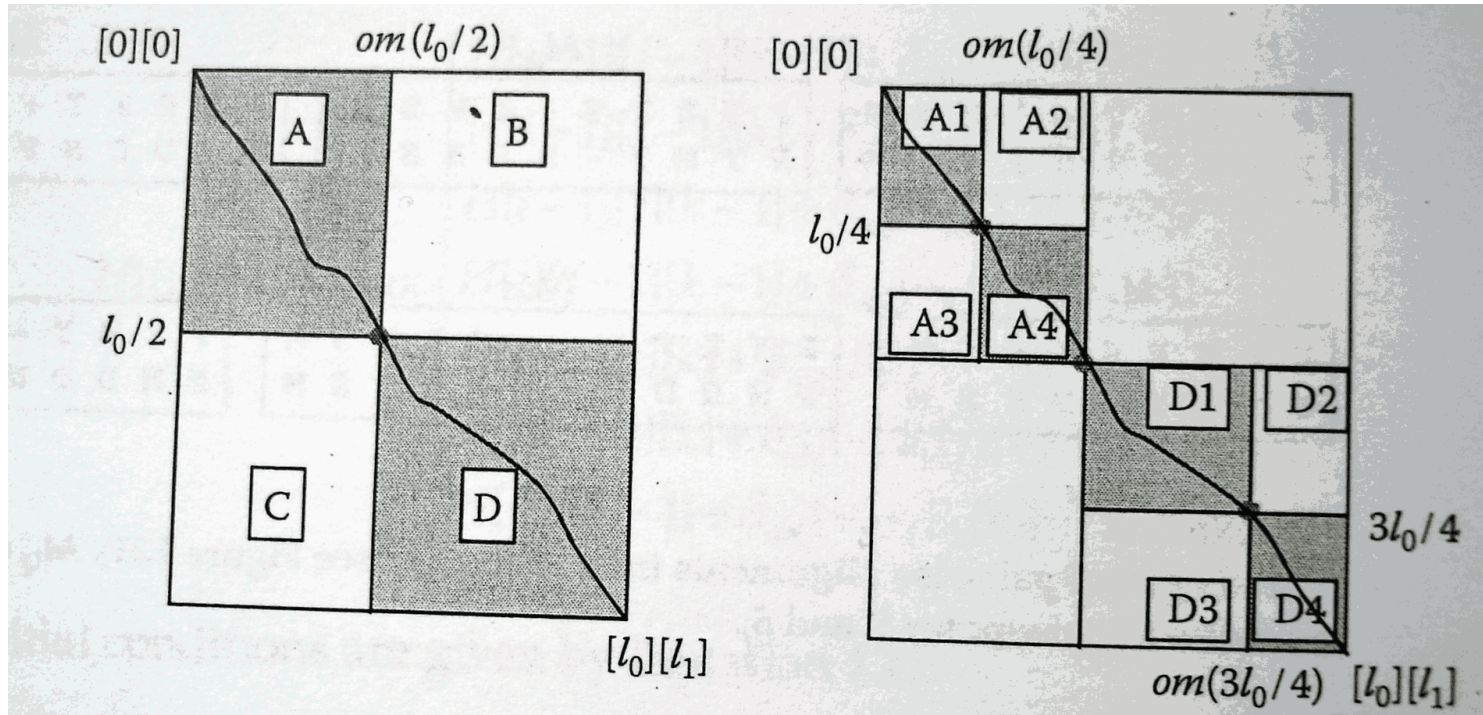


krok 1) przeprowadzamy dopasowanie dla górnej części, poczynając od punktu $[0,0]$, kończąc na wierszu $l_0/2$.

krok 2) przeprowadzamy dopasowanie dla dolnej części od tyłu (czyli dopasowujemy odwrócone sekwencje), poczynając od punktu $[l_0, l_0]$, kończąc na wierszu $l_0/2$.

krok 3) znajdujemy punktację *upper* dla górnej części macierzy i *lower* dla dolnej (sekwencja odwrócona). Sumujemy, znajdujemy punkt o najwyższej punktacji.

Poszukiwanie dopasowania - uściślenie



krok 1: cała macierz

krok 2: A

krok 3: A1

...

krok k: D

krok k+1: D1

...

Jak głęboko schodzimy z rekurencją?

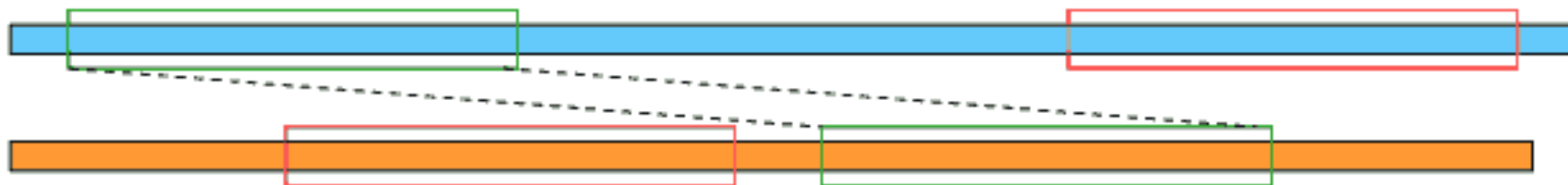
Do momentu aż rozmiar podmacierzy pozwoli nam na uruchomienie algorytmu w tradycyjnej formie.

Metody programowania dynamicznego mają zastosowanie dla porównywania par lub niewielkich zbiorów sekwencji.

W praktyce zwykle chcemy porównać naszą sekwencję ze wszystkimi sekwencjami zdeponowanymi w bazie – metoda staje się zbyt czasochłonna.

Dopasowania lokalne

Stosujemy dla sekwencji istotnie różniących się długością
(np. poszukiwanie krótkich motywów w całym genie
lub poszukiwanie genu w całym genomie)



Istota dopasowania lokalnego:
poszukiwanie regionów o dużym stopniu podobieństwa
(np. analiza białek wielodomenowych, poszukiwanie motywów)

Algorytm jest bardzo podobny do algorytmu N-W. Podobnie jak w N-W wykorzystujemy ideę programowania dynamicznego.

Różnica polega na tym, że w wypełnianej macierzy nie dopuszczamy wartości ujemnych.

$$H_{i,j} = \max \begin{cases} H_{i-1,j-1} + s(a_i, b_j) \\ H_{i-1,j} + g \\ H_{i,j-1} + g \\ 0 \end{cases}$$

H – macierz punktacji;

s – wartość dopasowania symboli a_i i b_j

g – kara za przerwę

Algorytm S-W: przykład (1)

Scoring Metric:

Match: $s(a_i, b_j) = 1$

Mismatch: $s(a_i, b_j) = -1$

Gap: -2 penalty

Maximum of possible scores:

(a) $0 + s(A,A) = 0 + 1 = 1$

(b) $0 - g = 0 - 2 = -2$

(c) $0 - g = 0 - 2 = -2$

(d) 0 (no pointer)

		A	A	T	G	T
		0	0	0	0	0
		0	0			
A		0	?			
T		0				
G		0				
A		0				
C		0				

Algorytm S-W: przykład (2)

Scoring Metric:

Match: $s(a_i, b_j) = 1$

Mismatch: $s(a_i, b_j) = -1$

Gap: -2 penalty

Maximum of possible scores:

(a) $0 + s(A,A) = 0 + 1 = 1$

(b) $1 - g = 1 - 2 = -1$

(c) $0 - g = 0 - 2 = -2$

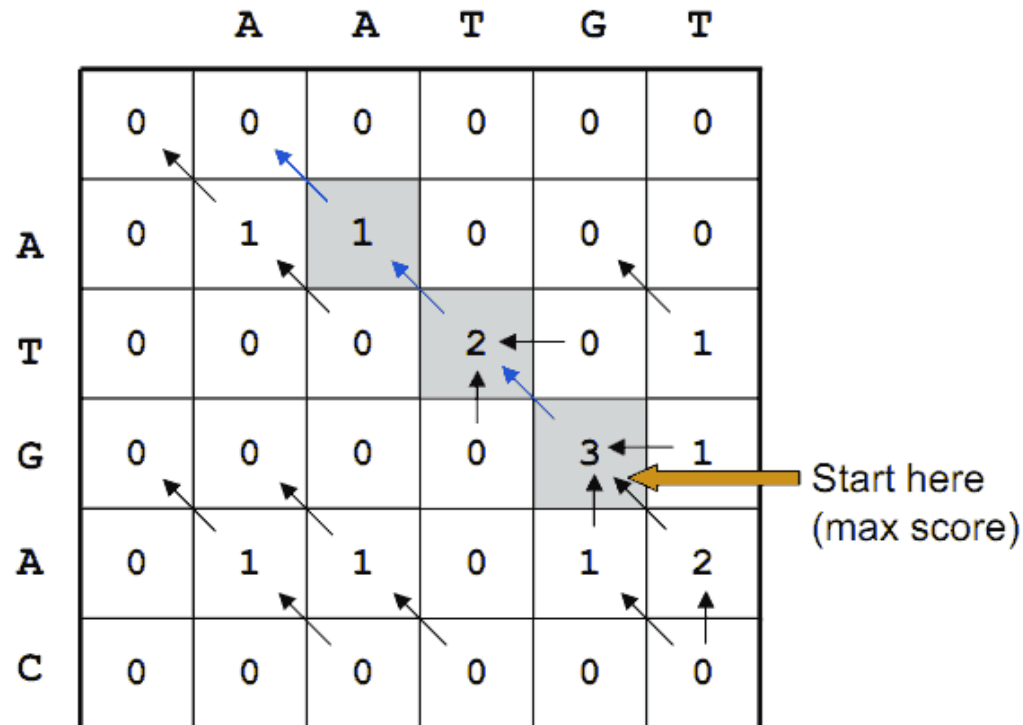
(d) 0 (no pointer)

		A	A	T	G	T
		0	0	0	0	0
A		0	1	?		
T		0				
G		0				
A		0				
C		0				

Diagram illustrating the Smith-Waterman algorithm grid. The grid shows scores for sequence alignment between 'A' (row) and 'A', 'A', 'T', 'G', 'T' (columns). The top-left cell (0,0) is 0. The cell (1,1) is 1, with arrows labeled 'a' (diagonal), 'b' (horizontal), and 'c' (vertical). The cell (1,2) is '?'. The cell (1,0) is 0. The cell (0,1) is 0. The cell (0,2) is 0. The cell (0,3) is 0. The cell (0,4) is 0. The cell (0,5) is 0. The cell (1,3) is 0. The cell (1,4) is 0. The cell (1,5) is 0. The cell (2,0) is 0. The cell (2,1) is 0. The cell (2,2) is 0. The cell (2,3) is 0. The cell (2,4) is 0. The cell (2,5) is 0. The cell (3,0) is 0. The cell (3,1) is 0. The cell (3,2) is 0. The cell (3,3) is 0. The cell (3,4) is 0. The cell (3,5) is 0. The cell (4,0) is 0. The cell (4,1) is 0. The cell (4,2) is 0. The cell (4,3) is 0. The cell (4,4) is 0. The cell (4,5) is 0. The cell (5,0) is 0. The cell (5,1) is 0. The cell (5,2) is 0. The cell (5,3) is 0. The cell (5,4) is 0. The cell (5,5) is 0.

Algorytm S-W: rekonstrukcja ścieżki

1. Startujemy z komórki o najwyższej wartości.
2. Kontynuujemy dopóki nie osiągniemy wartości 0.



Local Alignment AATGT
shown in blue: ATGAC

Optymalne dopasowania

Podobnie jak w algorytmie N-W, może istnieć wiele optymalnych dopasowań.

Mogą one mieć różną długość.

		A	E	K	P	C	A	Y	E	N	N	E	F
	0	0	0	0	0	0	0	0	0	0	0	0	0
E	0	0	1	0	0	0	0	0	1	0	0	1	0
K	0	0	0	2	1	0	0	0	0	0	0	0	0
C	0	0	0	1	1	2	1	0	0	0	0	0	0
A	0	1	0	0	0	1	3	2	1	0	0	0	0
Y	0	0	0	0	0	0	2	4	3	2	1	0	0
E	0	0	1	0	0	0	1	3	5	4	3	2	1
N	0	0	0	0	0	0	0	2	4	6	5	4	3
E	0	0	1	0	0	0	0	1	3	5	5	6	5
P	0	0	0	0	1	0	0	0	2	4	4	5	4
I	0	0	0	0	0	0	0	0	1	3	3	4	3
L	0	0	0	0	0	0	0	0	0	2	2	3	2
A	0	1	0	0	0	0	1	0	0	1	1	2	1

EKPCAYEN
EK-CAYEN

EKPCAYENNE
EK-CAYEN-E
EKPCAYENNE
EK-CAYE-NE

Tworzymy trzy macierze: H , E i F z warunkami brzegowymi:

$$\begin{aligned}
 H_{i,j} &= \max \begin{cases} H_{i-1,j-1} + s(a_i, b_j) \\ E_{i,j} \\ F_{i,j} \\ 0 \end{cases} & \begin{aligned} H_{0,0} &= 0 \\ H_{i,0} &= 0 \\ H_{0,j} &= 0 \end{aligned} \\
 E_{i,j} &= \max \begin{cases} H_{i-1,j} + Go + Ge \\ E_{i-1,j} + Ge \end{cases} & \begin{aligned} E_{0,0} &= -\infty \\ E_{i,0} &= Go + i * Ge \\ E_{0,j} &= -\infty \end{aligned} \\
 F_{i,j} &= \max \begin{cases} H_{i,j-1} + Go + Ge \\ F_{i,j-1} + Ge \end{cases} & \begin{aligned} F_{0,0} &= -\infty \\ F_{i,0} &= -\infty \\ F_{0,j} &= Go + j * Ge \end{aligned}
 \end{aligned}$$

How are the analysis coming?

Almost ready

