

CSE 182-L2: Blast & variants I

Dynamic Programming

www.cse.ucsd.edu/classes/fa09/cse182

www.cse.ucsd.edu/~vbafna

FA08

CSE182

Notes

- Assignment 1 is online, due next Tuesday.
- Discussion section is optional. Use it as a resource.
- On the web-site, you'll find some questions on lectures. Ideally, you should be able to answer the questions after attending these lectures (Not all of these are trivial, so please study them carefully).

FA08

CSE182

Searching Sequence databases

NCBI BLAST

10 February 2004 BLAST 2.2.8 has been released. Read more...

Nucleotide

- Discontiguous megablast
- Megablast
- Nucleotide-nucleotide BLAST (blastn)
- Search for short, nearly exact matches
- Search trace archives with megablast or discontiguous megablast

Protein

- Protein-protein BLAST (blastp)
- PHI- and PSI-BLAST
- Search for short, nearly exact matches
- Search the conserved domain database (psiblast)
- Search by domain architecture (cdart)

Translated

- Translated query vs. protein database (blastx)
- Protein query vs. translated database (tblastn)
- Translated query vs. translated database (tblastx)

Genomes

- Environmental samples
- Human, mouse, rat
- Fugu, zebrafish
- Insects, nematodes, plants, fungi, malaria
- Microbial genomes, other eukaryotic genomes

Special

- Align two sequences (bl2seq)
- Screen for vector contamination (VecScreen)
- Immunoglobulin BLAST (igblast)

Meta

- Retrieve results by RID
- Get this page with javascript-free links

[Disclaimer](#)
[Privacy statement](#)
[Accessibility](#)
Valid XHTML 1.0, CSS

<http://www.ncbi.nlm.nih.gov/BLAST/>

FA08

CSE182

Query:

```
>gi|26339572|dbj|BAC33457.1| unnamed protein product [Mus musculus]  
MSSTKLEDSLRRNWSSASELNETQEPFLNPTDYDDEEFLRYLWREYLHPKEYEYVLIAGYIIVFVVA  
LIGNVLVCVAVWKNHMRVTVNYFIVNLSLADVLVTTICLPATLVVDITETWFFGQSLCKVIPYLOTV  
SVSVSVLTLSCIALDRWYAICHPLMFKSTAKRARNISIVVIWIVSCIIMIPQAIVMCESSMLPGLANKT  
TLFTVCDEHWGGEVYPKMYHICFLVITYMAPLCLMILAYLQIFRKLWCRQIPGTSSVVQRKWKQQQP  
SQPRGSGQQSKARISAVAAEIKQIRARRKTARMLMVLLVFAICYLPISILNVLKRVFGMFTHTEDRE  
TVYAWFTFSHWLVIYANSAANPLIYNFLSGKFREEFKAAFSCCLGVHHRQGDRLARGRTSTESRKS  
LTTQISNFDNVSKLSEHVVLTSISTLPAANGAGPLQNWYLQQGVPSLLSTWLEV
```

- What is the function of this sequence?
- Is there a human homolog?
- Which cellular organelle does it work in? (Secreted/membrane bound)
- Idea: Search a database of known proteins to see if you can find similar sequences which have a known function

FA08

CSE182

Querying with Blast

The image displays two screenshots of the NCBI BLAST website. The left screenshot shows the BLAST Home page, which includes navigation tabs (Home, Recent Results, Saved Strategies, Help) and sections for BLAST Assembled Genomes, Basic BLAST, and Specialized BLAST. The right screenshot shows the BLAST search interface, which includes fields for 'Enter Query Sequence', 'Database' (set to Human genomic plus transcript), and 'Program Selection' (set to Highly similar sequences).

FA08

CSE182

Blast Output

- The output (Blastp query) is a series of protein sequences, ranked according to similarity with the query
- Each database hit is aligned to a subsequence of the query

FA08

CSE182

Blast Output 1

```

S Id → >gi|39930593|ref|NP_945197.1| hypocretin (orexin) receptor 1; orexin receptor
type 1 [F
] gi|37794009|gb|ARR01226.1| orexin receptor type-1
-----
Length = 416

Score = 526 bits (1354), Expect = e-148
Identities = 264/399 (66%), Positives = 307/399 (76%), Gaps = 14/399 (3%)

Q beg → Query: 26  E P F L N P T D Y D D E E P L A Y L W R E V L H P K R E Y E W V L I A G Y I I V F V V A L I G N V L V C V A W V W N H H M 85
      19  E P F P D S Y D E F L K Y L W R V L P K + Y E W V L I A Y + V F + A L + G N L V C + A V W + R H H M
Sbjct: 19  E P F H L P P Y E D E - E L K Y L W R D V L Y P K Q V E W V L I A Y V A V F L I A L V G N L V L V L A V W R H H M 77

Query: 86  R V T V N Y F I V N L S L A D V L V T T C L P A T L V - V D I T E T W F P G S L C R V I P Y L Q X X X X X X X X X X X X 145
Sbjct: 78  R V T V N Y F I V N L S L A D V L V T A I C L P A S L V D I T E S W L F G Q A L C R V I P Y L Q A V F V S V A V L E L 137

Query: 146  K C I A L D R W Y A I C H E L M P F S T A K R A R N S K X X X X X X X X X X X X X X X X X X X X X X X X X X X X X X X X X X X 205
Sbjct: 138  S P T A L D R W Y A I C H E L P K S T A R R A R G S I L G I W V S L A V H V P Q A A V M E C S S V L F E L A N R T R 197

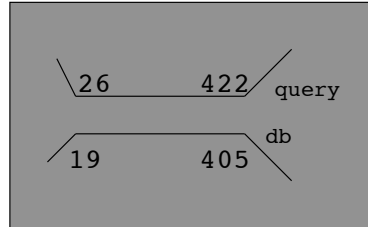
Query: 206  L P T V C D E H R G S E V Y P R M Y H I C F F L V T Y N A P L C M L I L A Y L Q I F R K L W C R Q I P C F S S V Q R K 265
Sbjct: 190  L E S V C D E H A D E L Y P K I Y H S C F T I V T Y L A P L G L M G H A Y F Q I F R K L W R Q I P C T S A L V N H 257

Query: 266  W E + + Q P V S Q P R G S G Q S K A I S A V A A E T K Q I R A R R K T A R N L M V V L L V F A I C Y L P I S I L 323
Sbjct: 258  W E R F S Q L H A Q H Q G L C T E F P P A R A F L A E V R Q M R A R R K T A R N L M V V L L V F A L C Y L P I S V L 317

Query: 324  N V L K R V F G M F T H T E D R E T V Y A W F T F S H W L V Y A N S A A N P I I Y N F L S G K F R E E P F A A F S C C L 393
Sbjct: 310  N V L K R V F G M F D R E V Y A F P F S H W L V Y A N S A A N P I I Y N F L S G K F R E + F A A F S C C L 377

Query: 384  G V H R Q C D R L A R G T S T E S R K S L T T Q T S S P N S V S E H 422
Sbjct: 378  P + + + + + S L C F G S S A R H R S L S L Q + + S R C S V S K V S E H 405
    
```

Schematic



FA08

CSE182

Blast Output 2 (drosophila)

```

S Id → >gi|45549233|ref|NP_524525.3| CG5811-PA [Drosophila melanogaster]
      gi|45446678|gb|AA035635.3| CG5811-PA [Drosophila melanogaster]
-----
Length = 464

Score = 137 bits (346), Expect = 5e-31
Identities = 89/316 (28%), Positives = 145/316 (45%), Gaps = 34/316 (10%)

Q beg → Query: 61  Y I I V F V V A L I G N V L V C V A W V W K N H H M R E V T H Y F I V N L S L A D V L V T T C L P A T L V - V D I T E T 119
      112 Y I P I F I F A L I G N C T V C Y I V Y S T P R M K T V T H Y F I A S L A I G D I L M S F F C V F S S F I S L F I L N Y 171
Sbjct: 112 Y I P I F I F A L I G N C T V C Y I V Y S T P R M K T V T H Y F I A S L A I G D I L M S F F C V F S S F I S L F I L N Y 171

Query: 120  W F F G Q S L C K V I P Y L Q X X X X X X X X X X X X X X X X X X X X X X X X X X X X X X X X X X X X X X X X X X X X X 179
Sbjct: 172  W F F G L A L C H F V N Y S Q A V S V L V S A Y T L V A I S I D R Y I A I M P L K P R I T R Y A T P I A G W W F I 231

Query: 180  X X X X X X P Q A I V M E C S S M L P G L A N K T L F T V C D E H W G G E V Y P R M Y H I C F F L V T Y N A P L C L M 239
Sbjct: 232  A L A T A L P I P I V S G L D I P M S P W H T K C E R Y - T C R E N W P S R T Q E Y Y T T L G L F A L Q V V P L Q V L 290

Query: 240  I L A Y L Q I F R E L N C R Q I P C T S S V V Q R K N K Q Q P V S Q P R G S G Q S K A I S A V A A E T K Q I R A R 299
Sbjct: 291  I F T Y A R I T R V M A K R P F G E A E T N R D Q + + + + + + + + + + + + + + + + + + + + + + + + + + + + + + + + + 322

Query: 300  R E T A R H L M V V L E V F A I C Y L P I S I L N V L K R V F G M F T H T E D R E T V Y A W F T F S H W L V Y A N S A A 359
Sbjct: 323  R E M V K H M L E V V I V F E C M L P F N I L Q L L N D - E E F A H W D - - P L P Y V W F A F - H W L A M S H C C Y 378

Query: 360  N P T I Y F L S G K F R E E F 375
Sbjct: 379  N P T I Y C M H A R E R G G + 394
    
```

FA08

CSE182

The technological question

- How do we measure similarity between sequences?
- Percent identity?

```
A T C A A C G      A T C A A - C G -  
T C A A T G G T    - T C A A T G G T
```

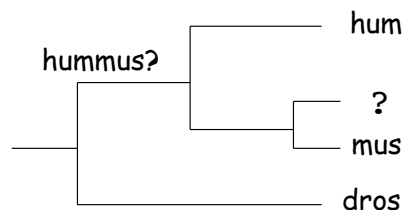
- Number of sequence edit operations?
 - Implies a notion of alignment that includes indels
- Technology question: Given two sequences, how do we compute a 'good' alignment? What is 'good'?

FA08

CSE182

The biology question

- How do we interpret these results?
 - Similar sequence in the 3 species implies that the common ancestor of the 3 had an ancestral form of that sequence.
 - The sequence accumulates mutations over time. These mutations may be indels, or substitutions.
- A 'good' alignment might be one in which many residues are identical. However,
 - Hum and mus diverged more recently and so the sequences are more likely to be similar.
 - Paralogs can create big problems



FA08

CSE182

Computing alignments

- What is an alignment?
 - $2 \times m$ table.
 - Each sequence is a row, with interspersed gaps
 - Columns describe the edit operations

A	A	-	T	C	G	G	A
	A	C	T	C	G	-	A

- What is the score of an alignment?
 - Score every column, and sum up the scores. Let C be the score function for the column
- How do we compute the alignment with the best score?

FA08

CSE182

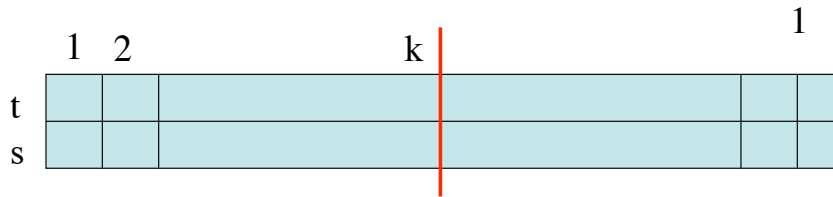
Optimum scoring alignments, and score of optimum alignment

- Instead of computing an optimum scoring alignment, we attempt to compute the score of an optimal alignment.
- Later, we will show that the two are equivalent

FA08

CSE182

Computing Optimal Alignment score

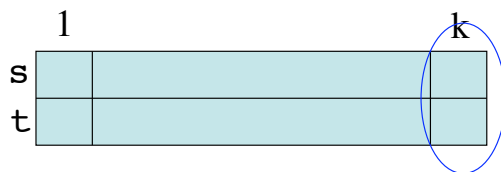


- Observations: The optimum alignment has nice recursive properties:
 - The alignment score is the sum of the scores of columns.
 - If we break off at cell k , the left part and right part must be optimal sub-alignments.
 - The left part contains prefixes $s[1..i]$, and $t[1..j]$ for some i and some j (we don't know the values of i and j).

FA08

CSE182

Optimum prefix alignments



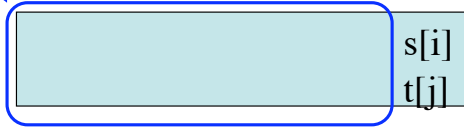
- Consider an optimum alignment of the prefix $s[1..i]$, and $t[1..j]$
- Look at the last cell, indexed by k . It can only have 3 possibilities.

FA08

CSE182

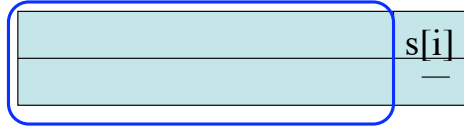
3 possibilities for rightmost cell

Optimum alignment of $s[1..i-1]$, and $t[1..j-1]$

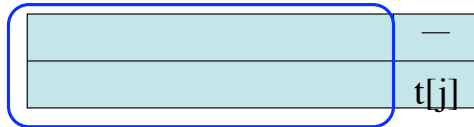


1. $s[i]$ is aligned to $t[j]$

Optimum alignment of $s[1..i-1]$, and $t[1..j]$



2. $s[i]$ is aligned to '-'



3. $t[j]$ is aligned to '-'

Optimum alignment of $s[1..i]$, and $t[1..j-1]$

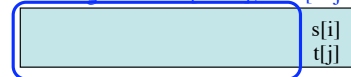
FA08

CSE182

Optimal score of an alignment

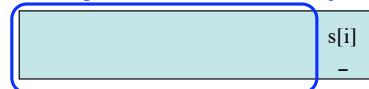
$$S[i,j] = C(s_i, t_j) + S(i-1, j-1)$$

Optimum alignment of $s[1..i-1]$, and $t[1..j-1]$



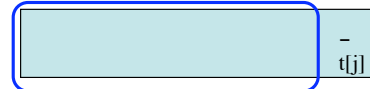
$$S[i,j] = C(s_i, -) + S(i-1, j)$$

Optimum alignment of $s[1..i-1]$, and $t[1..j]$



$$S[i,j] = C(-, t_j) + S(i, j-1)$$

Optimum alignment of $s[1..i]$, and $t[1..j-1]$



- Let $S[i,j]$ be the score of an optimal alignment of the prefix $s[1..i]$, and $t[1..j]$. It must be one of 3 possibilities.

Optimal alignment score

$$S[i, j] = \max \begin{cases} S[i-1, j-1] + C(s_i, t_j) \\ S[i-1, j] + C(s_i, -) \\ S[i, j-1] + C(-, t_j) \end{cases}$$

- Which prefix pairs (i,j) should we use? For now, simply use all.
- If the strings are of length m, and n, respectively, what is the score of the optimal alignment?

FA08

CSE182

Sequence Alignment

- Recall: Instead of computing the optimum alignment, we are computing the score of the optimum alignment
- Let $S[i, j]$ denote the score of the optimum alignment of the prefix $s[1..i]$ and $t[1..j]$

FA08

CSE182

An $O(nm)$ algorithm for score computation

For $i = 1$ to n

For $j = 1$ to m

$$S[i,j] = \max \begin{cases} S[i-1,j-1] + C(s_i,t_j) \\ S[i-1,j] + C(s_i,-) \\ S[i,j-1] + C(-,t_j) \end{cases}$$

- The iteration ensures that all values on the right are computed in earlier steps.

FA08

CSE182

Base case (Initialization)

$$S[0,0] = 0$$

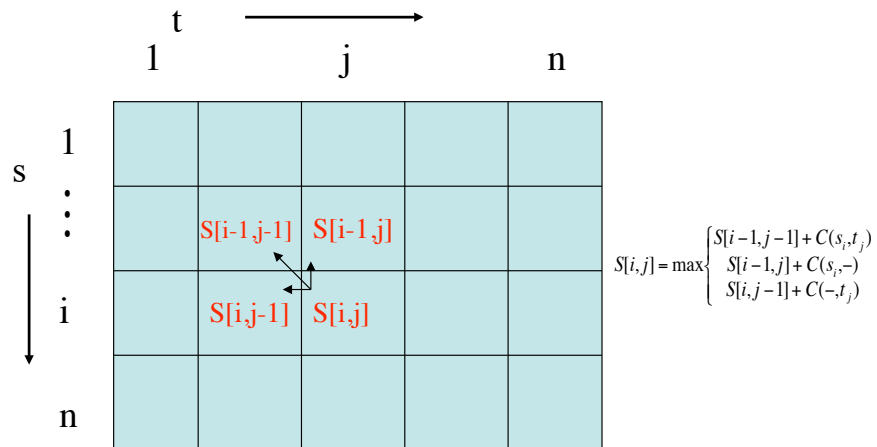
$$S[i,0] = C(s_i,-) + S[i-1,0] \quad \forall i$$

$$S[0,j] = C(-,s_j) + S[0,j-1] \quad \forall j$$

FA08

CSE182

A tableaux approach



Cell (i,j) contains the score $S[i,j]$. Each cell only looks at 3 neighboring cells

FA08

CSE182

An Example

T	C	A	T	-
T	G	C	A	A

A1

T	C	A	T
T	G	C	A

A2

- Align $s=TCAT$ with $t=TGCAA$
- Match Score = 1
- Mismatch score = -1, Indel Score = -1
- Score A1?, Score A2?

FA08

CSE182

Alignment Table

	T	G	C	A	A	
T	0	-1	-2	-3	-4	-5
C	-1	1	0	-1	-2	-3
A	-2	0	0	1	0	-1
T	-3	-1	-1	0	2	1
T	-4	-2	-2	-1	1	1

FA08

CSE182

Alignment Table

	T	G	C	A	A	
T	0	-1	-2	-3	-4	-5
C	-1	1	0	-1	-2	-3
A	-2	0	0	1	0	-1
T	-3	-1	-1	0	2	1
T	-4	-2	-2	-1	1	1

- $S[4,5] = 1$ is the score of an optimum alignment
- Therefore, A2 is an optimum alignment
- We know how to obtain the optimum Score. How do we get the best alignment?

FA08

CSE182

Computing Optimum Alignment

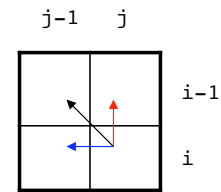
- At each cell, we have 3 choices
- We maintain additional information to record the choice at each step.

For $i = 1$ to n

For $j = 1$ to m

$$S[i,j] = \max \begin{cases} S[i-1,j-1] + C(s_i, t_j) \\ S[i-1,j] + C(s_i, -) \\ S[i,j-1] + C(-, t_j) \end{cases}$$

If $(S[i,j] = S[i-1,j-1] + C(s_i, t_j))$ $M[i,j] =$ ↖
 If $(S[i,j] = S[i-1,j] + C(s_i, -))$ $M[i,j] =$ ↑
 If $(S[i,j] = S[i,j-1] + C(-, t_j))$ $M[i,j] =$ ←



FA07

CSE182

Computing Optimal Alignments

		T	G	C	A	A
	0	←-1	←-2	←-3	←-4	←-5
T	↑-1	↖1	←0	←-1	←-2	←-3
C	↑-2	↑0	↖0	↖1	←0	←-1
A	↑-3	↑-1	↖-1	↑0	↖2	↖1
T	↑-4	↑-2	↖-2	↑-1	↑1	↖1

FA07

CSE182

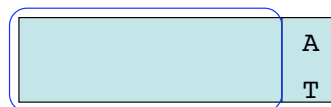
Retrieving Opt.Alignment

		1	2	3	4	5
		T	G	C	A	A
1	T	0	-1	-2	-3	-4
2	C	-1	1	0	-1	-2
3	A	-2	0	0	1	0
4	T	-3	-1	-1	0	2
		-4	-2	-2	-1	1

• $M[4,5]=$

Implies that $S[4,5]=S[3,4]+C(A,T)$

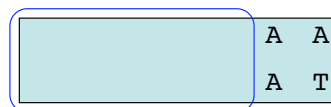
or



$M[3,4]=$

Implies that $S[3,4]=S[2,3]+C(A,A)$

or



FA07

CSE182

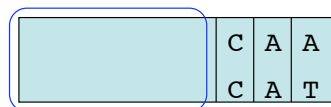
Retrieving Opt.Alignment

		1	2	3	4	5
		T	G	C	A	A
1	T	0	-1	-2	-3	-4
2	C	-1	1	0	-1	-2
3	A	-2	0	0	1	0
4	T	-3	-1	-1	0	2
		-4	-2	-2	-1	1

• $M[2,3]=$

Implies that $S[2,3]=S[1,2]+C(C,C)$

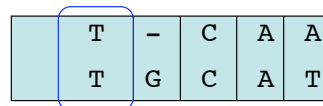
or



$M[1,2]=$

Implies that $S[1,2]=S[1,1]+C(-,G)$

or



FA08

CSE182

Algorithm to retrieve optimal alignment

RetrieveAl(i,j)

if (M[i,j] == ` \ `)

s _i
t _j

)

else if (M[i,j] == ` | `)

s _i
-

)

else if (M[i,j] == ` -- `)

-
t _j

)

FA08

CSE182

Summary

- An optimal alignment of strings of length n and m can be computed in $O(nm)$ time
- There is a tight connection between computation of optimal score, and computation of opt. Alignment
 - True for all DP based solutions

FA08

CSE182

Global versus Local Alignment

Consider

s = ACCACCCCTT

t = ATCCCACAT

```

ACCACCCCTT
| | | | |
ATCCCACAT
    
```

```

ACCACCCCT T
| | | | |
A TCCCACAT
    
```

Sometimes, this is preferable

FA08

CSE182

Blast Outputs Local Alignment

```

>gi|39930993|ref|NP_045197.1| hypocretin (orexin) receptor 1; orexin receptor
type 1 [Mus
musculus] gi|37704009|gb|ARR01326.1| orexin receptor type-1 [Mus
musculus]
Length = 416

Score = 526 bits (1354), Expect = e-148
Identities = 264/399 (66%), Positives = 307/399 (76%), Gaps = 14/399 (3%)

Query: 26  ERFLENPTDYDDEEFLRYLWREYLPKREYEWVLIAGYIIIVFVVALIGVLCVAVWNRHM 85
Sbjct: 19  ERF P DY+DE FLRYLWR+YL+PK+YEWVLEIA Y+ VF++AL+GN LVC+AVW+NRHM 77

Query: 86  RVTNFFIVNLSLADVLVTTCLPATELVVDITETWFGQSLCRVIFYLQXXXXXXXXXX 145
Sbjct: 78  RVTNFFIVNLSLADVLVTAICLPAELVDITEHW FGQ+LCRVIFYLQ 137

Query: 146  KCIALDRWYAIChELMFSFAKRRANSXXXXXXXXXXXXPQAVMECSHLPGLAKRT 205
Sbjct: 138  SPTALDRWYAIChELPSTARRRAGSILGINAVSLAVMVPQAAMVECSVLPGLAKRTR 197

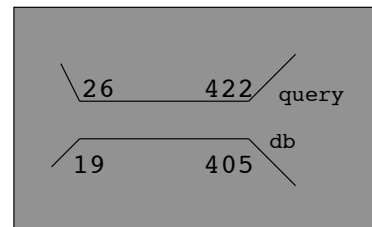
Query: 206  LFTVCDENHGSEVVPKMYHICFFELVYMAPLCEMLIAYLQIFRKLWCRQIPGTSVVQRK 245
Sbjct: 190  LFTVCDENHDELFPKMYHICFFELVYMAPLCEMLIAYLQIFRKLWCRQIPGTSVAVR 257

Query: 266  WQ--QPVSQPRGSGQSKARTSAVAARTKQIRARRKTAHMLVLLVFAICYLPISEIL 323
Sbjct: 250  WRPSEQLHAHQGLCTEPQRRARAFLEAVVQHRARRKTAHMLVLLVFAICYLPISEVL 317

Query: 324  NVLKRVFCHFTHTEDRETVYAWPTFPHWLVYANSANPITVYVFLSGKFRFPAAFSCCL 383
Sbjct: 318  NVLKRVFCHFTDRVYVAFPTFPHWLVYANSANPITVYVFLSGKFRFPAAFSCCL 377

Query: 384  GVHHRQCRLARGRTSTESRSLSLTQISNFDNSELSEH 422
Sbjct: 378  P-----GLPGSARRKSLSLQ--SRCSVSRVSEH 405
    
```

Schematic

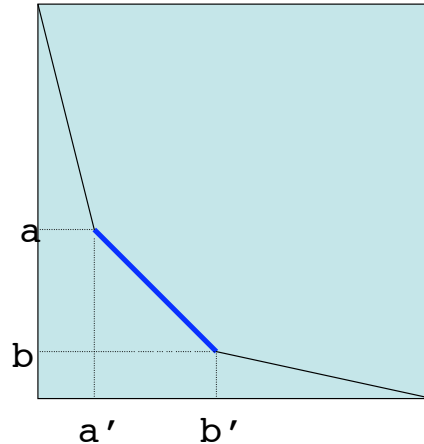


FA08

CSE182

Local Alignment

- Compute maximum scoring interval over all sub-intervals (a,b) , and (a',b')
- How can we compute this efficiently?

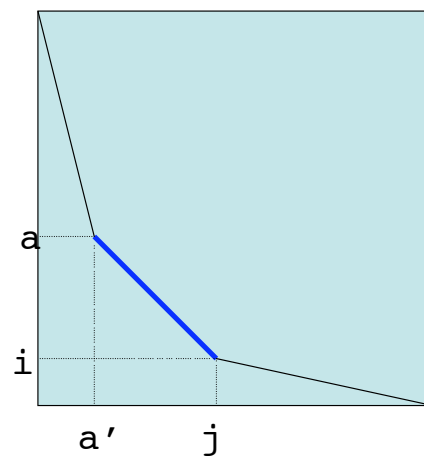


FA08

CSE182

Local Alignment

- Recall that in global alignment, we compute the best score for all prefix pairs $s(1..i)$ and $t(1..j)$.
- Instead, compute the best score for all sub-alignments that end in $s(i)$ and $t(j)$.
- What changes in the recurrence?

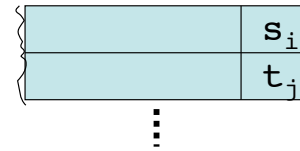
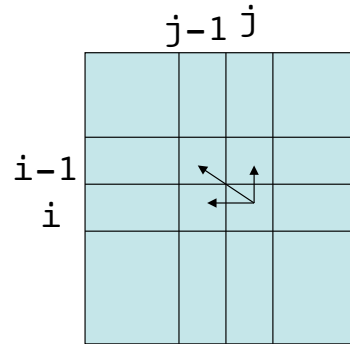


FA08

CSE182

Local Alignment

- The original recurrence still works, except when the optimum score $S[i,j]$ is negative
- When $S[i,j] < 0$, it means that the optimum local alignment cannot include the point (i,j) .
- So, we must reset the score to 0.



FA08

CSE182

Local Alignment Trick (Smith-Waterman algorithm)

$$S[i,j] = \max \begin{cases} 0 \\ S[i-1, j-1] + C(s_i, t_j) \\ S[i-1, j] + C(s_i, -) \\ S[i, j-1] + C(-, t_j) \end{cases}$$

How can we compute the local alignment itself?

FA08

CSE182

Generalizing Gap Cost

- It is more likely for gaps to be contiguous
- The penalty for a gap of length l should be

$$go + ge * l$$

```
>gi|39930593|ref|NP_945197.1| hypocretin (orexin) receptor 1; orexin receptor
type 1 [Mus
musculus] gi|37794009|gb|AA01326.1| orexin receptor type-1 [Mus
musculus]
Length = 416
Score = 526 bits (1354), Expect = e-148
Identities = 264/399 (66%), Positives = 307/399 (76%), Gaps = 14/399 (3%)
Query: 26  EPELNPTDYDDEEFLRYLWREYLNPREYENVLIAQYIIIVPVALIGNVLVCVAVWKRHHN 85
EPF P DY+DE FLRYLWR+YL+PE+YENVLIA Y+ VP++AL+GN LVC+AVW+RHHN
Sbjct: 19  EPEHLFPDYEDS-FLRYLWRDYLFPQYENVLIAAVVAVFLIALVGNLVLCLAVWRHHN 77
Query: 86  RTVTNRFIVHLSLADVLVTTTCLPAPLAVDITETWFFGSLCKVIFPLQXXXXXXXXXXXX 145
RTVTNRFIVHLSLADVLVTF CLEA+L+VDITE+W FGQ+LCKVIFPLQ
Sbjct: 78  RTVTNRFIVHLSLADVLVTAICLPASLLVDITESEVFGQALCKVIFPLQAVSVSVAVLTL 137
Query: 146  KCTALDRNYAICHPEMPEFSTARRRANSXXXXXXXXXXXXPQAIWRECSNLPGLANRTT 205
IALDRNYAICHPEL+FKSTM+RAR S PQA WRECS+LD LAN+P
Sbjct: 138  SPTIALDRNYAICHPELFRSTMARRRANSILGTMVSLAVRFPQAVWRECS+VFLERANTR 197
Query: 206  LPTVCDENHGGEVYPMYHICFFLAVTYRNPICLMLAYLQIFRKLWCRQIPGTSVVQRK 265
LF+VCDENH H+YPK+YH CFF+VTY+APL LM +AY QIFRKLW RQIPGT+S + R
Sbjct: 198  LPSVCDENHADBLPKIYHSCFFIVTYLAPLGLKSHAYFQIFRKLWGRQIPGTTSAVRRN 257
Query: 266  WEQ--QQPVSQPRGSGQSEKARTSAAVAETKQIRARRRTARRHLMVLLVFAICVLPISIL 323
WE+ +Q +G + + R A AB+KQ+RARRRTA+HLMVLLVFA+CYLPIS+L
Sbjct: 258  WERPSEGLEAQQGLCTHFPQPARAFLEAVKQHRARRRTARRHLMVLLVFAICVLPISVL 317
Query: 324  NVEKRVFGHFTHTDRETFVYAMFTFSSHLLVYANSAANP IITYNPLSGKFRBPKAAFSCL 393
NVEKRVFGHF DRH VFA FTFSSHLLVYANSAANP IITYNPLSGKFRB+PKAAFSCL
Sbjct: 318  NVEKRVFGHFRAQSDRENVACTFTFSSHLLVYANSAANP IITYNPLSGKFRBQPKAAFSCL 377
Query: 384  GVHHRQCDRLARGRTSTESRKSILTQIENFDNVKLSKH 422
G S+ KSL+ Q + +VSK+SEH
Sbjct: 378  P-----SLCPGSSARRKSLSLQ--SRCSVSKVSEH 405
```

FA07

CSE182

End of Lecture 2

FA07

CSE182