CSE 417: Algorithms and Computational Complexity

W. L. Ruzzo

Dynamic Programming, II
RNA Folding
Outline

A few (well, ~30) slides on applications of dynamic programming in biology (not on exams or anything, but you might enjoy a slightly deeper look at the application of some of the algorithms we study)

- Sequence alignment
- RNA structure

Algorithms for RNA structure
(yes, this part is fodder for hw & exams)
Application: Sequence Search
Moore’s Law
Growth of GenBank (Base Pairs)


Excludes “short-read archive,” > 7 terabases by mid-2009
SRA database growth

2,492,489,806,780,758 total bases
1,174,010,897,059,076 open access bases

http://www.ncbi.nlm.nih.gov/Traces/sra/i/g.png
Sequencing Costs Outpace Moore’s Law
A Database Search

go to, e.g., [http://www.uniprot.org/](http://www.uniprot.org/), “blast” tab, and paste in this:

```plaintext
>sp|P15172|MYOD1_HUMAN Myoblast determination protein 1 OS=Homo sapiens GN=MYOD1 PE=1 SV=3
MELLSPPLRDVLTDAPDGLCSFATDDFYDDPCFDSPDLPFFEDLDPRMLHMVGALLKPE
EHSHFPAAVHPAPGAREDEHVRAPSGHHQAGRCLLLWACKACKRKTINADRRKAATMRERR
RLSKVNEAFETLKRCCTSSNPQRPLKVEILRNAYYEGLQALLRQDAAPPGAAAAFYA
PGPLLPPRRGGEHYSGDSDASSPRSNCSDGMDYSGPPSOGRRNCYEGAYYNEAPSEP
PGKSAAVSSLIDCLSSIVERISTESPAPALLLADVSPSPPRRQEAAPSEGESSGDPTQS
PDAAPQQCPAGANPNPIYQVL
```
A Few seconds Later…

...And 100’s more…

<table>
<thead>
<tr>
<th>Accession</th>
<th>Entry name</th>
<th>0Query hit320</th>
<th>0Match hit (sqrt scale)</th>
<th>Name (Organism)</th>
</tr>
</thead>
<tbody>
<tr>
<td>P15172</td>
<td>MYOD1_HUMAN</td>
<td></td>
<td></td>
<td>human Myoblast determination protein 1 (Homo sapiens)</td>
</tr>
<tr>
<td>B2RC72</td>
<td>B2RC72_HUMAN</td>
<td></td>
<td></td>
<td>human cDNA, FLJ95884, highly similar to Hom… (Homo sapiens)</td>
</tr>
<tr>
<td>E2RT59</td>
<td>E2RT59_CANFA</td>
<td></td>
<td></td>
<td>dog Uncharacterized protein (Canis familiaris)</td>
</tr>
<tr>
<td>P49811</td>
<td>MYOD1_PIG</td>
<td></td>
<td></td>
<td>pig Myoblast determination protein 1 (Sus scrofa)</td>
</tr>
<tr>
<td>D2KPI9</td>
<td>D2KPI9_PIG</td>
<td></td>
<td></td>
<td>pig Myogenic differentiation 1 (Sus scrofa)</td>
</tr>
<tr>
<td>F1S9A9</td>
<td>F1S9A9_PIG</td>
<td></td>
<td></td>
<td>pig Uncharacterized protein (Sus scrofa)</td>
</tr>
<tr>
<td>D2IOV4</td>
<td>D2IOV4_AILME</td>
<td></td>
<td></td>
<td>panda Putative uncharacterized protein (Ailuropoda melanoleuca)</td>
</tr>
<tr>
<td>P29331</td>
<td>MYOD1_SHEEP</td>
<td></td>
<td></td>
<td>sheep Myoblast determination protein 1 (Ovis aries)</td>
</tr>
<tr>
<td>D2SP11</td>
<td>D2SP11_BUBBU</td>
<td></td>
<td></td>
<td>water buffalo Myogenic factor MYOD1 (Bubalus bubalis)</td>
</tr>
<tr>
<td>Q0VBX9</td>
<td>Q0VBX9_BOVIN</td>
<td></td>
<td></td>
<td>cow Myogenic differentiation 1 (Bos taurus)</td>
</tr>
<tr>
<td>Q7YS82</td>
<td>MYOD1_BOVIN</td>
<td></td>
<td></td>
<td>cow Myoblast determination protein 1 (Bos taurus)</td>
</tr>
<tr>
<td>Q8C6B1</td>
<td>Q8C6B1_MOUSE</td>
<td></td>
<td></td>
<td>mouse Myogenic differentiation 1 (Mus musculus)</td>
</tr>
<tr>
<td>A0JPK9</td>
<td>A0JPK9_RAT</td>
<td></td>
<td></td>
<td>rat Myogenic differentiation 1 (Rattus norvegicus)</td>
</tr>
<tr>
<td>Q02346</td>
<td>MYOD1_RAT</td>
<td></td>
<td></td>
<td>rat Myoblast determination protein 1 (Rattus norvegicus)</td>
</tr>
<tr>
<td>P10085</td>
<td>MYOD1_MOUSE</td>
<td></td>
<td></td>
<td>mouse Myoblast determination protein 1 (Mus musculus)</td>
</tr>
<tr>
<td>Q6DTY5</td>
<td>Q6DTY5_PIG</td>
<td></td>
<td></td>
<td>pig Eukaryotic myogenic factor MYF-3 (Sus scrofa)</td>
</tr>
<tr>
<td>P21572</td>
<td>MYOD1_COTJA</td>
<td></td>
<td></td>
<td>quail Myoblast determination protein 1 homolog (Columbidae coturnix japonica)</td>
</tr>
<tr>
<td>Q6DV59</td>
<td>Q6DV59_MELGA</td>
<td></td>
<td></td>
<td>turkey MyoD (Meleagris gallopavo)</td>
</tr>
<tr>
<td>P16075</td>
<td>MYOD1_CHICK</td>
<td></td>
<td></td>
<td>chicken Myoblast determination protein 1 homolog (Gallus gallus)</td>
</tr>
<tr>
<td>C5J072</td>
<td>C5J072_CHICK</td>
<td></td>
<td></td>
<td>chicken Myogenic differentiation 1 (Gallus gallus)</td>
</tr>
<tr>
<td>C3U011</td>
<td>C3U011_ANAPL</td>
<td></td>
<td></td>
<td>chicken Myogenic differentiation 1 (Anas platyrhynchos)</td>
</tr>
<tr>
<td>F1NHM3</td>
<td>F1NHM3_CHICK</td>
<td></td>
<td></td>
<td>chicken Uncharacterized protein (Gallus gallus)</td>
</tr>
<tr>
<td>F1NXM5</td>
<td>F1NXM5_CHICK</td>
<td></td>
<td></td>
<td>chicken Uncharacterized protein (Gallus gallus)</td>
</tr>
<tr>
<td>P13904</td>
<td>MYODA_XENLA</td>
<td></td>
<td></td>
<td>frog Myoblast determination protein 1 homolog A (Xenopus laevis)</td>
</tr>
<tr>
<td>Q6AVZ0</td>
<td>Q6AVZ0_XENLA</td>
<td></td>
<td></td>
<td>frog Myod1-a protein (Xenopus laevis)</td>
</tr>
<tr>
<td>Q7T109</td>
<td>Q7T109_XENTR</td>
<td></td>
<td></td>
<td>MyoD protein (Xenopus tropicalis)</td>
</tr>
<tr>
<td>Accession</td>
<td>Entry name</td>
<td>Status</td>
<td>Protein names</td>
<td>Organism</td>
</tr>
<tr>
<td>-----------</td>
<td>------------------</td>
<td>--------</td>
<td>---------------</td>
<td>-----------------------------------------------</td>
</tr>
<tr>
<td>Q7T109</td>
<td>Q7T109_XENTR</td>
<td>✭</td>
<td>MyoD protein</td>
<td><em>Xenopus tropicalis (Western clawed frog)</em></td>
</tr>
</tbody>
</table>

**Alignment 1 against Q7T109**

<table>
<thead>
<tr>
<th>Score</th>
<th>E-value</th>
<th>Identity</th>
<th>Query length</th>
<th>Match length</th>
<th>Position</th>
</tr>
</thead>
<tbody>
<tr>
<td>964</td>
<td>1.0 x 10^{-102}</td>
<td>64.0%</td>
<td>320</td>
<td>74.0%</td>
<td>288</td>
</tr>
</tbody>
</table>

*Q7T109 matches from 1 to 288 (288AA), in the query sequence from 1 to 320 (320AA)*

**Graphical Alignment**

```
1  MELLSPPLRLDVLTDAPGSGLCSFATTDDFYDPCFDSPDLRFFEDLDPLRMHVAGALLKPE  60  P15172
   MELL PLRD+++T +GSLCSF T DDFYDPCF++ D+ FFEDLDPRLVHALLKPE
1  MELLPPPLRLMEVT--EGSLSCLSDKPDDFYDPCFDNRLDSFFEDLDPRLVHALLKPE  57  Q7T109

61  EHSHPAAVHPAPGAREDHEHRPAGSHHQAAGRCLLLWACKACKRRTKTNDRRKAAATMRERR  120  P15172
+ H  EDEHRPAGSHHQAAGRCLLLWACKACKRRTKTNDRRKAAATMRERR
58  DPHH--------NEDEHRPAGSHHQAAGRCLLLWACKACKRRTKTNDRRKAAATMRERR  106  Q7T109

121  RLSKVNEAFETLKRCTSSNPQRLPKVEILRNAIYIEGLQALLRDQDAAPGGAAAFYA  180  P15172
     RLSKVNEAFETLKRCTS+PNQRLPKVEILRNAIYIE LQ+LLR Q+ FY
107  RLSKVNEAFETLKRCTSTNPQRLPKVEILRNAIYESLQSSLRQGE--------EFY-  158  Q7T109

181  PGGLPPGREGHEYGDSGDSASPRSNCSDMYDGSPPSGARRNCYEYGNAYNAPSEPRP  240  P15172
     P+ EYSGDSGDSASPRSNCSDGDMYSPP+RRNINGY+ ++Y+++P+ R
159  --PVL------EYSGDSGDSASPRSNCSGDMDYSPPCGRRNNSYDSFYSIDSPNGLRL  210  Q7T109

241  GKSAAVSSLMLCSSIERFSTPAAPALLADVPSESPQRPQREEAAPSEGSS-----SGDP  297  P15172
     GKS+ +SSLMLCSSIERFST P + AD SE P +P +GE+ SG
211  GKSVISSLMLCSSIERFSTPVCPVIPAAADSGSEGSP-----CSPLQGETLSEGII  265  Q7T109
```
hits at rank ~ 250 still extremely good matches, even though very distantly related organisms (and rank 1000+…)

<table>
<thead>
<tr>
<th>Filter</th>
<th>Overview</th>
<th>Results</th>
<th>Customize order</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alignments</td>
<td>Entry</td>
<td>Entry name</td>
<td>Status</td>
</tr>
<tr>
<td>B3LY60</td>
<td>GF18746</td>
<td>Drosophila ananassae (Fruit fly)</td>
<td>334</td>
</tr>
<tr>
<td>O4RJ8</td>
<td>Q4RGJ8</td>
<td>Chromosome undetermined SCAFF099, whole genome</td>
<td>126</td>
</tr>
<tr>
<td>F1NF6</td>
<td>F1NF6_CHECH</td>
<td>Gallus gallus (Chicken)</td>
<td>237</td>
</tr>
<tr>
<td>Q91151</td>
<td>Q91151_NOTVI</td>
<td>Myogenic regulatory factor; transcription factor</td>
<td>219</td>
</tr>
<tr>
<td>Q29BN7</td>
<td>Q29BN7_DROS</td>
<td>Drosophila pseudoobscura pseudoobscura (Fruit fly)</td>
<td>330</td>
</tr>
<tr>
<td>B4G81</td>
<td>GL13832</td>
<td>Drosophila persimilis (Fruit fly)</td>
<td>330</td>
</tr>
<tr>
<td>Q92020</td>
<td>MYF6_XENLA</td>
<td>Myogenic factor 6</td>
<td>240</td>
</tr>
<tr>
<td>B7ZGB0</td>
<td>B7ZGB0_XENLA</td>
<td>Xenopus laevis (African clawed frog)</td>
<td>240</td>
</tr>
<tr>
<td>A7UC1</td>
<td>A7UC1_XENLA</td>
<td>Xenopus laevis (African clawed frog)</td>
<td>240</td>
</tr>
<tr>
<td>F7F8X</td>
<td>F7F8X_MONDO</td>
<td>Myogenin</td>
<td>243</td>
</tr>
<tr>
<td>D9IV56</td>
<td>D9IV56_EPIDO</td>
<td>Epinephelus cooidec (Orange-spotted grouper)</td>
<td>250</td>
</tr>
<tr>
<td>Q6SYV5</td>
<td>MYF6_TAKRU</td>
<td>Myogenic factor 6</td>
<td>239</td>
</tr>
<tr>
<td>G3JP0</td>
<td>G3JP0_SARHA</td>
<td>Sarcoptes scabiei (Tasmanian devil)</td>
<td>243</td>
</tr>
<tr>
<td>G7Y4S</td>
<td>G7Y4S_SSLUIS</td>
<td>Transcription factor SUM-1</td>
<td>946</td>
</tr>
<tr>
<td>G4LX1</td>
<td>G4LX1_SCHMA</td>
<td>Myogenic factor, putative</td>
<td>864</td>
</tr>
<tr>
<td>G1EN3</td>
<td>G1EN3_DUGA</td>
<td>Dugesia japonica (Planarian)</td>
<td>498</td>
</tr>
<tr>
<td>Q8W4P</td>
<td>Q8W4P_SCHMD</td>
<td>MyoD-like protein</td>
<td>466</td>
</tr>
<tr>
<td>Q6PV5</td>
<td>MYF6_TETNG</td>
<td>Myogenic factor 6</td>
<td>239</td>
</tr>
<tr>
<td>Q2RZ12</td>
<td>Q2RZ12_SALSA</td>
<td>Myogenin</td>
<td>254</td>
</tr>
<tr>
<td>Q92078</td>
<td>Q92078_ONCNY</td>
<td>TM2ogenin protein</td>
<td>254</td>
</tr>
</tbody>
</table>
The foregoing search capability is a *huge* deal

the “google” of molecular biology

millions of searches daily

biologists (not just “computational” biologists) use this routinely

it connects information about *all* living things

(dynamic programming)

Time permitting, more on algorithm later …
Application: RNA structure
The Double Helix

As shown, the two strands coil about each other in a fashion such that all the bases project inward toward the helix axis. The two strands are held together by hydrogen bonds (pink rods) linking each base projecting from one backbone to its so-called complementary base projecting from the other backbone. The base A always bonds to T (A and T are comple-

Shown in (b) is an uncoiled fragment of (a). Each three complementary base pair is composed of a polymer made up of four related compounds called deoxyribonucleotides.

http://www.rcsb.org/pdb/explore.do?structureId=1GAT
Fig. 2. The arrows show the situation as it seemed in 1958. Solid arrows represent probable transfers, dotted arrows possible transfers. The absent arrows (compare Fig. 1) represent the impossible transfers postulated by the central dogma. They are the three possible arrows starting from protein.
Non-coding RNA

Messenger RNA - codes for proteins
Non-coding RNA - all the rest

Before, say, mid 1990’s, 1-2 dozen known
(critically important, but narrow roles: e.g., tRNA)

Since mid 90’s dramatic discoveries

Regulation, transport, stability/degradation
E.g. “miRNA”: >1000 in humans; regulate >50% of genes
E.g. “riboswitches”: 10000’s in bacteria

By some estimates, ncRNA >> mRNA
DNA structure: dull

5’...ACCGCTAGATG...3’

3’...TGGCGATCTAC...5’
RNA Secondary Structure:
RNA makes helices too

Base pairs
A=U
C=G

Usually single stranded
RNA Secondary Structure:

Not everything, but important, easier than 3d
Why is structure important?

• For protein-coding, similarity in sequence is a powerful tool for finding related sequences
  – e.g. “hemoglobin,” “MyoD” and many others are easily recognized in all animals

• For many non-coding RNAs, different sequences can have the same structure, and structure is most important for function.
  – So, using structure plus sequence, can find related sequences at much greater evolutionary distances
6S mimics an open promoter

Barrick et al. RNA 2005
Trotochaud et al. NSMB 2005
Willkomm et al. NAR 2005
In Bacteria: A typical biosynthetic cycle around a critical metabolite ("SAM")
Gene Regulation: The MET Repressor

Protein

DNA

Alberts, et al. 3e.
Not the only way!

Protein way  Riboswitch alternative

Epshtein, et al., PNAS 2003
Winkler et al., Nat. Struct. Biol. 2003
Not the only way!

Protein way

Riboswitch alternatives

SAM-II


Corbino et al., Genome Biol. 2005
Not the only way!

Protein way

Riboswitch alternatives

SAM-III

Corbino et al., Genome Biol. 2005

Fuchs et al., NSMB 2006

Alberts, et al., 3e.

Corbino et al., Gen. Biol. 2005

SAM-I

SAM-II
Not the only way!

Protein way

Riboswitch alternatives

Corbino et al., Genome Biol. 2005
Fuchs et al., NSMB 2006
Weinberg et al., RNA 2008
Alberts, et al., 3e.

Corbino et al., Genome Biol. 2005

Fuchs et al., NSMB 2006

Weinberg et al., RNA 2008


SAM-I

SAM-II

SAM-III

SAM-IV

SAM-V

Protein way

Riboswitch alternatives

Not the only way!
And many other examples. Widespread, deeply conserved, structurally sophisticated, functionally diverse, biologically important uses for ncRNA throughout prokaryotic world.

boxed = confirmed riboswitch (+2 more)

Why is RNA hard to deal with?

A: Structure often more important than sequence \(^{31}\)
Origin of Life?

Life needs
information carrier: DNA
molecular machines, like enzymes: Protein
making proteins needs DNA + RNA + proteins
making (duplicating) DNA needs proteins
Horrible circularities! How could it have arisen in an abiotic environment?
Origin of Life?

RNA can carry information, too
  RNA double helix; RNA-directed RNA polymerase
RNA can form complex structures
RNA enzymes exist (ribozymes)
RNA can control, do logic (riboswitches)

The “RNA world” hypothesis:
  1st life was RNA-based
6.5 RNA Secondary Structure

Nussinov’s Algorithm – core technology for RNA structure prediction
RNA Secondary Structure

RNA. String $B = b_1 b_2 \ldots b_n$ over alphabet \{ A, C, G, U \}.

Secondary structure. RNA is usually single-stranded, and tends to loop back and form base pairs with itself. This structure is essential for understanding molecular behavior.

Ex: GUCGAUUGAGCGAAUGUAACCAACGUGGCUACGGCGAGA
RNA Secondary Structure (somewhat oversimplified)

Secondary structure. A set of pairs $S = \{ (b_i, b_j) \}$ that satisfy:

- [Watson-Crick.]
  - $S$ is a matching, i.e. each base pairs with at most one other, and
  - each pair in $S$ is a Watson-Crick pair: A-U, U-A, C-G, or G-C.

- [No sharp turns.] The ends of each pair are separated by at least 4 intervening bases. If $(b_i, b_j) \in S$, then $i < j - 4$.

- [Non-crossing.] If $(b_i, b_j)$ and $(b_k, b_l)$ are two pairs in $S$, then we cannot have $i < k < j < l$. (Violation of this is called a pseudoknot.)

Free energy. Usual hypothesis is that an RNA molecule will form the structure with the optimum total free energy.

Goal. Given an RNA molecule $B = b_1b_2...b_n$, find a secondary structure $S$ that maximizes the number of base pairs.

approximated by maximizing number of base pairs
RNA Secondary Structure: Examples

Examples.

- **base pair**

- **ok**

- **sharp turn**

- **crossing**
RNA Secondary Structure: Subproblems

First attempt. \( \text{OPT}[j] = \) maximum number of base pairs in a secondary structure of the substring \( b_1b_2 \ldots b_j \).

Difficulty. Results in two sub-problems.

- Finding secondary structure in: \( b_1b_2 \ldots b_{t-1} \) \( \leftarrow \text{OPT}(t-1) \)
- Finding secondary structure in: \( b_{t+1}b_{t+2} \ldots b_{j-1} \) \( \leftarrow \) not “OPT” of anything; need more flexible set of sub-problems
Dynamic Programming Over Intervals: (R. Nussinov’s algorithm)

Notation. \( \text{OPT}[i, j] = \text{maximum number of base pairs in a secondary structure of the substring } b_ib_{i+1}\ldots b_j. \)

- **Case 1.** If \( i \geq j - 4. \)
  \[
  \text{OPT}[i, j] = 0 \text{ by no-sharp turns condition.}
  \]

- **Case 2.** Base \( b_j \) is not involved in a pair.
  \[
  \text{OPT}[i, j] = \text{OPT}[i, j-1]
  \]

- **Case 3.** Base \( b_j \) pairs with \( b_t \) for some \( i \leq t < j - 4. \)
  Non-crossing constraint decouples resulting sub-problems
  \[
  \text{OPT}[i, j] = 1 + \max_t \{ \text{OPT}[i, t-1] + \text{OPT}[t+1, j-1] \}
  \]
  take max over \( t \) such that \( i \leq t < j-4 \) and \( b_t \) and \( b_j \) are Watson-Crick complements

**Key point:** Either last base is unpaired (case 1, 2) or paired (case 3)

**Remark.** Core idea in CKY algorithm for context-free parsing
“Optimal pairing of $b_i \ldots b_j$”

Two possibilities:

j Unpaired:
Find best pairing of $b_i \ldots b_{j-1}$

j Paired (with some $t$):
Find best $b_i \ldots b_{t-1}$ +
best $b_{t+1} \ldots b_{j-1}$ plus 1

Why is it slow?
Why do pseudoknots matter?
Bottom Up Dynamic Programming Over Intervals

Q. What order to solve the sub-problems?
A. One way—do shortest intervals first:

```plaintext
RNA(b_1, ..., b_n) {
    for k = 5, 6, ..., n-1
        for i = 1, 2, ..., n-k
            j = i + k
            Compute OPT[i, j]
            using recurrence
    return OPT[1, n]
}
```

Running time. $O(n^3)$. 
Nussinov: Max Pairing

\[ \text{opt}(i,j) = \# \text{ pairs in optimal pairing of } b_i \ldots b_j \]
\[ \text{opt}(i,j) = 0 \text{ for all } i, j \text{ with } i \geq j-4; \text{ otherwise} \]
\[ \text{opt}(i,j) = \text{max of:} \]
\[ \begin{align*}
\quad \text{opt}(i,j-1) \\
\quad \max \{ \text{opt}(i,t-1)+1+\text{opt}(t+1,j-1) \mid \\
\quad \quad i \leq t < j-4 \text{ and } b_t-b_j \text{ may pair} \}
\end{align*} \]

Another Computation Order

\[ \text{opt}(i,j) = \# \text{ pairs in optimal pairing of } b_i \ldots b_j \]

\[ \text{opt}(i,j) = 0 \text{ for all } i, j \text{ with } i \geq j-4; \]

otherwise:

\[ \text{opt}(i,j) = \max \text{ of:} \]

\[ \left\{ \begin{array}{l}
\text{opt}(i,j-1) \\
\max \left\{ \text{opt}(i,t-1)+1+\text{opt}(t+1,j-1) \mid i \leq t < j-4 \text{ and } b_t-b_j \text{ may pair} \right\}
\end{array} \right. \]

Time: \( O(n^3) \)
Which Pairs?

Usual dynamic programming “trace-back” tells you which base pairs are in the optimal solution, not just how many.

Details? : homework
(Examples here and below assume 1-based indexing)
Computing one cell: \( \text{OPT}[2, 18] = ? \)

\[
\text{OPT}(i, j) = \begin{cases} 
0 & \text{if } i \geq j - 4 \\
\max \left\{ \text{OPT}[i, j - 1], 1 + \max_{t} (\text{OPT}[i, t - 1] + \text{OPT}[t + 1, j - 1]) \right\} & \text{otherwise}
\end{cases}
\]

Case 1:
2 ≥ 18-4? no.

Case 2:
\( B_{18} \) unpaired?
Always a possibility; then \( \text{OPT}[2, 18] \geq 3 \)

\[ \text{GGAAAACCCCAAAGGGGU} \]
\[ ((\ldots))\, (\ldots) \ldots \]
Computing one cell: $\text{OPT}[2,18] = ?$

$$\text{OPT}(i,j) = \begin{cases} 
0 & \text{if } i \geq j - 4 \\
\max \left\{ \text{OPT}[i, j - 1], 1 + \max_t (\text{OPT}[i, t - 1] + \text{OPT}[t + 1, j - 1]) \right\} & \text{otherwise}
\end{cases}$$

Case 3, $2 \leq t < 18 - 4$: $t = 2$: no pair
Computing one cell: \( \text{OPT}[2,18] = ? \)

\[
\text{OPT}(i, j) = \begin{cases} 
0 & \text{if } i \geq j - 4 \\
\max \left\{ \begin{array}{c} 
\text{OPT}[i, j - 1] \\
1 + \max_t (\text{OPT}[i, t - 1] + \text{OPT}[t + 1, j - 1])
\end{array} \right. & \text{otherwise}
\end{cases}
\]

Case 3, \( 2 \leq t < 18-4: \) t = 3: no pair
Computing one cell: $\text{OPT}[2,18] = ?$

$$\text{OPT}(i,j) = \begin{cases} 0 & \text{if } i \geq j - 4 \\ \max \left\{ \text{OPT}[i, j - 1], 1 + \max_t (\text{OPT}[i, t - 1] + \text{OPT}[t + 1, j - 1]) \right\} & \text{otherwise} \end{cases}$$

Case 3, $2 \leq t < 18 - 4$:
$t = 4$: yes pair
$\text{OPT}[2,18] \geq 1 + 0 + 3$

GGAAACCCAAAGGGGU
\ldots (\ldots (((((\ldots))))))
Computing one cell: $\text{OPT}[2, 18] = ?$

|   | G | G | G | A | A | A | C | C | C | A | A | A | G | G | G | G | U | U | U |
| 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 2 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 4 | 5 | 6 |
| 2 | 0 | 0 | 0 | 0 | 0 | 1 | 2 | 2 | 2 | 2 | 2 | 2 | 3 | 3 | 3 | 4 | 5 | 6 |
| 3 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 2 | 2 | 3 | 3 | 4 | 5 | 6 |
| 4 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 2 | 2 | 3 | 4 | 5 | 6 |
| 5 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 2 | 2 | 3 | 4 | 5 | 6 |
| 6 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 2 | 2 | 3 | 4 | 5 | 6 |
| 7 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 2 | 2 | 3 | 4 | 5 | 6 |
| 8 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 2 | 2 | 3 | 4 | 4 | 4 |
| 9 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 2 | 2 | 3 | 3 | 3 | 3 |
| 10 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 2 | 2 | 3 | 3 | 3 | 3 |
| 11 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 2 | 2 | 2 | 2 | 3 |
| 12 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 2 | 3 |
| 13 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 2 | 3 |
| 14 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| 15 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 16 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 17 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 18 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

Case 3, $2 \leq t < 18 - 4$:

$t = 5$: yes pair

$\text{OPT}[2, 18] \geq 1 + 0 + 3$

$\text{OPT}(i,j) = \begin{cases} 0 & \text{if } i \geq j - 4 \\ \max \left\{ \begin{array}{l} \text{OPT}[i, j - 1] \\ 1 + \max_i (\text{OPT}[i, t - 1] + \text{OPT}[t + 1, j - 1]) \end{array} \right\} & \text{otherwise} \end{cases}$
Computing one cell: $\text{OPT}[2,18] = ?$

Case 3, $2 \leq t < 18 - 4$:

$t = 6$: yes pair

$\text{OPT}[2,18] \geq 1 + 0 + 3$

$\text{GGAAACCCAAAGGGGU}$

\[ \ldots ((((((\ldots)))))) \]

Computing one cell:

\[
\text{OPT}(i, j) = \begin{cases} 
0 & \text{if } i \geq j - 4 \\
\max \left\{ \begin{array}{c} 
\text{OPT}[i, j - 1] \\
1 + \max_t (\text{OPT}[i, t - 1] + \text{OPT}[t + 1, j - 1]) 
\end{array} \right. & \text{otherwise}
\end{cases}
\]
Computing one cell: $OPT[2,18] = \ ?$

$$OPT(i,j) = \begin{cases} 
0 & \text{if } i \geq j - 4 \\
\max \left\{ \begin{array}{l}
OPT[i,j - 1] \\
1 + \max_t (OPT[i,t - 1] + OPT[t + 1, j - 1])
\end{array} \right\} & \text{otherwise}
\end{cases}$$

Case 3, $2 \leq t < 18 - 4$:
$\begin{array}{c}
t = 7: \text{yes pair} \\
OPT[2,18] \geq 1 + 0 + 3
\end{array}$

$GGAAAACCCAAAGGGGU$

$\ldots \ldots ( ( ( ( \ldots ) ) ) )$
Computing one cell: \( \text{OPT}[2, 18] = ? \)

\[
\begin{align*}
\text{Case 3, } 2 \leq t < 18-4: \\
t = 8: \text{no pair}
\end{align*}
\]

\[
\text{OPT}(i, j) = \begin{cases} 
0 & \text{if } i \geq j - 4 \\
\max \left\{ \text{OPT}[i, j - 1], 1 + \max_t \left( \text{OPT}[i, t - 1] + \text{OPT}[t + 1, j - 1] \right) \right\} & \text{otherwise}
\end{cases}
\]
Computing one cell: $\text{OPT}[2,18] = ?$

$\text{OPT}(i,j) = \begin{cases} 
0 & \text{if } i \geq j - 4 \\
\max \left\{ \text{OPT}[i,j-1] + 1 + \max_t (\text{OPT}[i,t-1] + \text{OPT}[t+1,j-1]) \right\} & \text{otherwise}
\end{cases}$

Case 3, $2 \leq t < 18-4$: $t = 11$: yes pair

$\text{OPT}[2,18] \geq 1 + 2 + 0$

$\text{GGAAAACCCAAAGGGGU}$

$(\ldots)$ $(\ldots)$

(not shown: $t = 9, 10, 12, 13$)
Overall, Max = 4 several ways, e.g.:

GGAAACCCAAAGGGGU

..(...(((....))))

tree shows trace back:
square = case 3
octagon = case 1

Computing one cell: 
OPT[2,18] = 4

\[
OPT(i,j) = \begin{cases} 
0 & \text{if } i \geq j - 4 \\
\max \left\{ \left\{ OPT[i,j - 1] \right\}, 1 + \max_t (OPT[i,t - 1] + OPT[t + 1, j - 1]) \right\} & \text{otherwise}
\end{cases}
\]
A  L19 (rplS) mRNA leader

<table>
<thead>
<tr>
<th>35</th>
<th>10</th>
<th>P1</th>
<th>P2</th>
<th>RBS</th>
<th>Start</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bsu</td>
<td>TGCAT</td>
<td>AAGAUGCUCCUGGCUCCGCGCGCUUUUGC...UAGGAGCAUG.05</td>
<td>AAGGAGCAUG.08</td>
<td>AUG</td>
<td></td>
</tr>
<tr>
<td>Bha</td>
<td>CTCAT</td>
<td>AAGAUGCUCCUGGCUCCGCGCGCUUUUGC...UAGGAGCAUG.06</td>
<td>AAGGAGCAUG.11</td>
<td>AUG</td>
<td></td>
</tr>
<tr>
<td>Gho</td>
<td>TTTTT</td>
<td>AAGAUGCUCCUGGCUCCGCGCGCUUUUGC...UAGGAGCAUG.04</td>
<td>AAGGAGCAUG.07</td>
<td>AUG</td>
<td></td>
</tr>
<tr>
<td>Gka</td>
<td>TTTCT</td>
<td>AAGAUGCUCCUGGCUCCGCGCGCUUUUGC...UAGGAGCAUG.08</td>
<td>AAGGAGCAUG.12</td>
<td>AUG</td>
<td></td>
</tr>
<tr>
<td>Bcl</td>
<td>TTTCT</td>
<td>AAGAUGCUCCUGGCUCCGCGCGCUUUUGC...UAGGAGCAUG.06</td>
<td>AAGGAGCAUG.12</td>
<td>AUG</td>
<td></td>
</tr>
<tr>
<td>Bac</td>
<td>GTTAC</td>
<td>AAGAUGCUCCUGGCUCCGCGCGCUUUUGC...UAGGAGCAUG.06</td>
<td>AAGGAGCAUG.07</td>
<td>AUG</td>
<td></td>
</tr>
<tr>
<td>Lmo</td>
<td>CAAATC</td>
<td>AAGAUGCUCCUGGCUCCGCGCGCUUUUGC...UAGGAGCAUG.04</td>
<td>AAGGAGCAUG.05</td>
<td>AUG</td>
<td></td>
</tr>
<tr>
<td>Sau</td>
<td>ATCAGA</td>
<td>AAGAUGCUCCUGGCUCCGCGCGCUUUUGC...UAGGAGCAUG.04</td>
<td>AAGGAGCAUG.05</td>
<td>AUG</td>
<td></td>
</tr>
<tr>
<td>Cpe</td>
<td>TTTAA</td>
<td>08GUCCGCCCGAUGUGCAGCCAA...AGAGUGCUAAG.17</td>
<td>AAGGAGCAUG.11</td>
<td>AUG</td>
<td></td>
</tr>
<tr>
<td>Chy</td>
<td>TTTCT</td>
<td>AAGAUGCUCCUGGCUCCGCGCGCUUUUGC...UAGGAGCAUG.03</td>
<td>AAGGAGCAUG.09</td>
<td>AUG</td>
<td></td>
</tr>
<tr>
<td>Swo</td>
<td>AAAAA</td>
<td>AAGAUGCUCCUGGCUCCGCGCGCUUUUGC...UAGGAGCAUG.05</td>
<td>AAGGAGCAUG.07</td>
<td>AUG</td>
<td></td>
</tr>
<tr>
<td>Ame</td>
<td>CCACCA</td>
<td>AAGAUGCUCCUGGCUCCGCGCGCUUUUGC...UAGGAGCAUG.07</td>
<td>AAGGAGCAUG.07</td>
<td>AUG</td>
<td></td>
</tr>
<tr>
<td>Dre</td>
<td>CAAA</td>
<td>AAGAUGCUCCUGGCUCCGCGCGCUUUUGC...UAGGAGCAUG.04</td>
<td>AAGGAGCAUG.12</td>
<td>AUG</td>
<td></td>
</tr>
<tr>
<td>Spn</td>
<td>ACACAG</td>
<td>AAGAUGCUCCUGGCUCCGCGCGCUUUUGC...UAGGAGCAUG.06</td>
<td>AAGGAGCAUG.07</td>
<td>AUG</td>
<td></td>
</tr>
<tr>
<td>Smu</td>
<td>ATACG</td>
<td>AAGAUGCUCCUGGCUCCGCGCGCUUUUGC...UAGGAGCAUG.04</td>
<td>AAGGAGCAUG.05</td>
<td>AUG</td>
<td></td>
</tr>
<tr>
<td>Lpl</td>
<td>ATTAC</td>
<td>AAGAUGCUCCUGGCUCCGCGCGCUUUUGC...UAGGAGCAUG.04</td>
<td>AAGGAGCAUG.05</td>
<td>AUG</td>
<td></td>
</tr>
<tr>
<td>Efa</td>
<td>TTTAA</td>
<td>08GUCCGCCCGAUGUGCAGCCAA...AGAGUGCUAAG.17</td>
<td>AAGGAGCAUG.11</td>
<td>AUG</td>
<td></td>
</tr>
<tr>
<td>Ljo</td>
<td>TAGAC</td>
<td>AAGAUGCUCCUGGCUCCGCGCGCUUUUGC...UAGGAGCAUG.06</td>
<td>AAGGAGCAUG.08</td>
<td>AUG</td>
<td></td>
</tr>
<tr>
<td>Sth</td>
<td>AGAGA</td>
<td>AAGAUGCUCCUGGCUCCGCGCGCUUUUGC...UAGGAGCAUG.03</td>
<td>AAGGAGCAUG.07</td>
<td>AUG</td>
<td></td>
</tr>
<tr>
<td>Lac</td>
<td>TTTCG</td>
<td>AAGAUGCUCCUGGCUCCGCGCGCUUUUGC...UAGGAGCAUG.03</td>
<td>AAGGAGCAUG.08</td>
<td>AUG</td>
<td></td>
</tr>
<tr>
<td>Spj</td>
<td>TTAGA</td>
<td>AAGAUGCUCCUGGCUCCGCGCGCUUUUGC...UAGGAGCAUG.06</td>
<td>AAGGAGCAUG.07</td>
<td>AUG</td>
<td></td>
</tr>
<tr>
<td>Lsa</td>
<td>TTTAA</td>
<td>08GUCCGCCCGAUGUGCAGCCAA...AGAGUGCUAAG.17</td>
<td>AAGGAGCAUG.11</td>
<td>AUG</td>
<td></td>
</tr>
<tr>
<td>Lsl</td>
<td>TTTCG</td>
<td>AAGAUGCUCCUGGCUCCGCGCGCUUUUGC...UAGGAGCAUG.04</td>
<td>AAGGAGCAUG.07</td>
<td>AUG</td>
<td></td>
</tr>
<tr>
<td>Fnu</td>
<td>TTAGA</td>
<td>AAGAUGCUCCUGGCUCCGCGCGCUUUUGC...UAGGAGCAUG.04</td>
<td>AAGGAGCAUG.12</td>
<td>AUG</td>
<td></td>
</tr>
</tbody>
</table>

Example: Ribosomal Autoregulation

Excess L19 represses L19 (RF00556; 555-559 similar)

B. subtilis L19 mRNA leader
Covariation is strong evidence for base pairing
Summary

RNA has important roles
   Beyond mRNA; many unexpected recent discoveries
Structure is critical to function
   True of other molecules, too
RNA secondary structure prediction is a key tool
Dynamic programming—useful accuracy, $O(n^3)$ time:
   Binary choice again: last base is paired or not
   Optimal substructure again: given last pair, optimally fold inside & outside separately
   Tabulate again: best folding of all substrings.