CSE 421: Intro Algorithms

W. L. Ruzzo

Dynamic Programming:
String alignment and RNA Folding

Outline

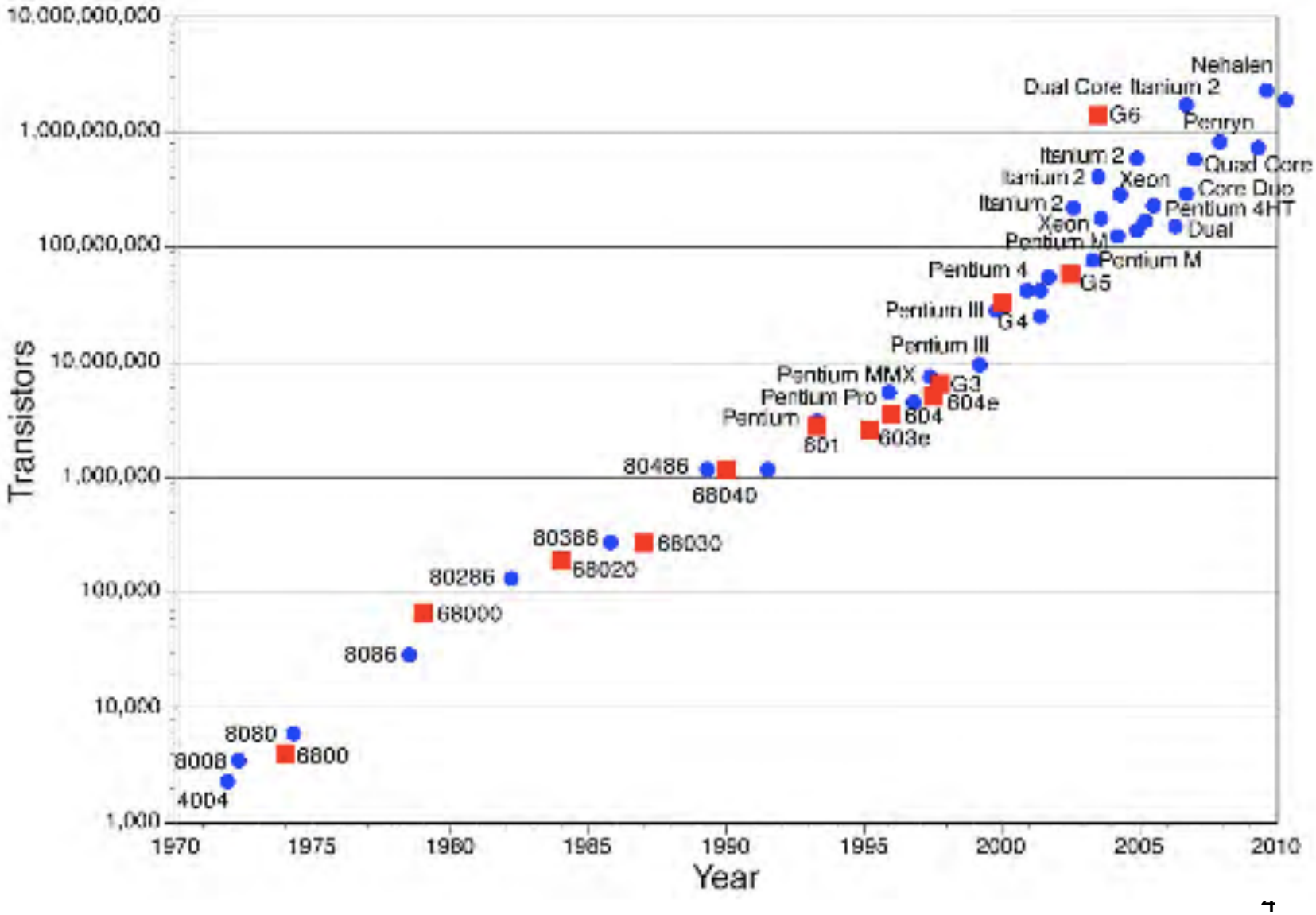
A few slides on *applications* of dynamic programming in biology

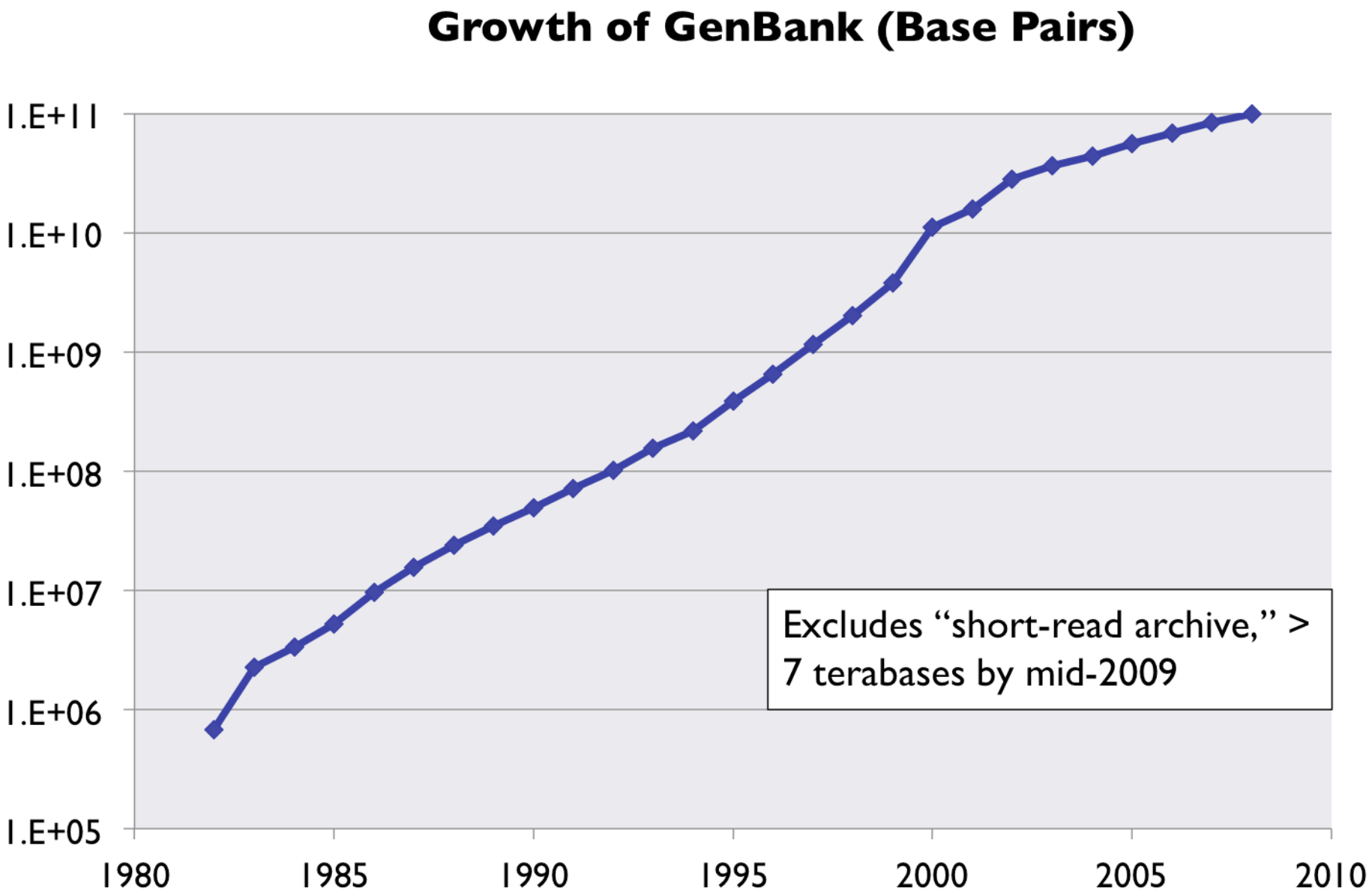
- Sequence alignment
- RNA structure

Algorithms for RNA structure

Application: Sequence Search

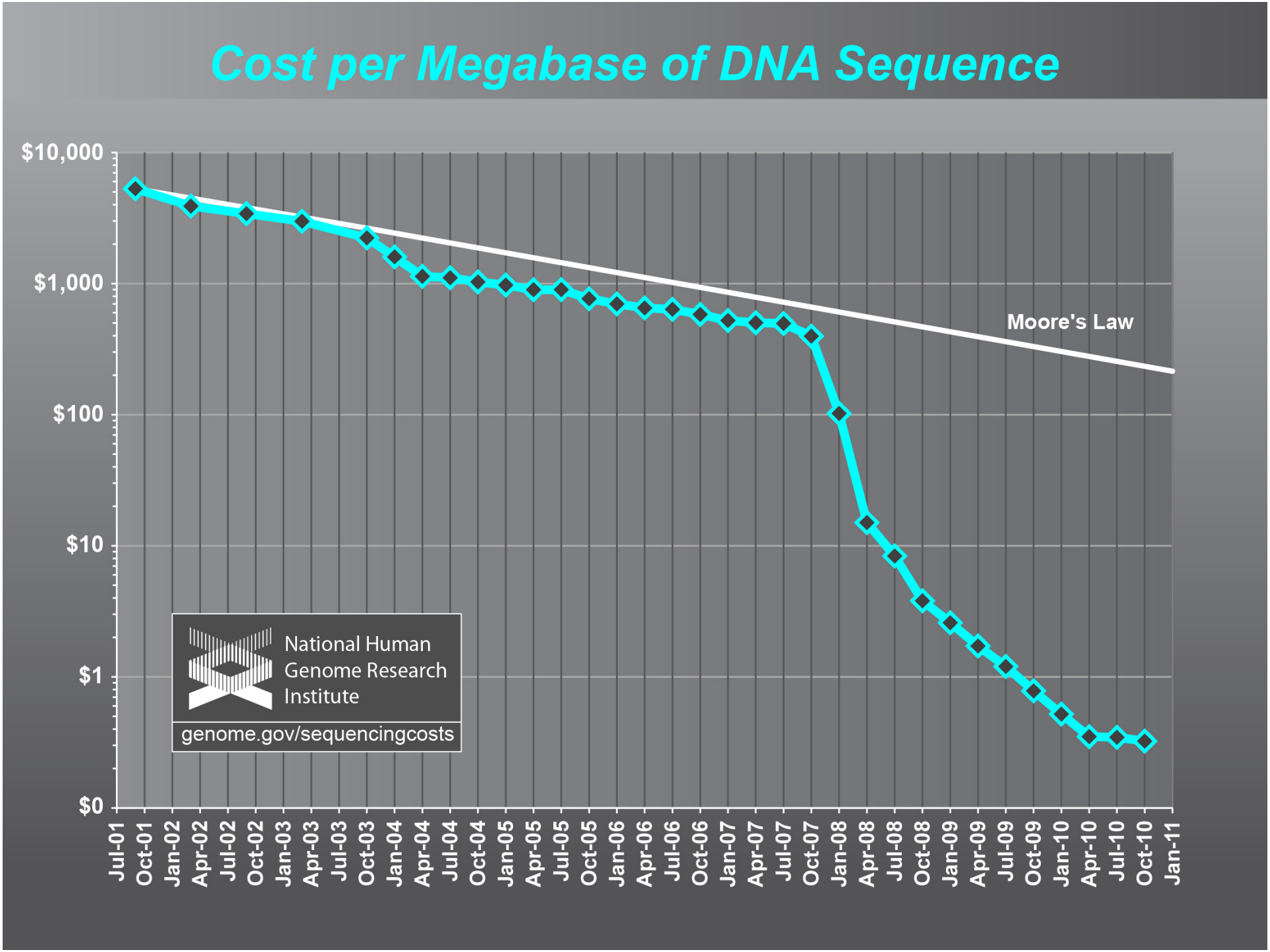
Moore's Law


10,000,000,000
1,000,000,000
100,000,000
10,000,000
1,000,000
100,000
10,000
1,000
Transistors
1970
1975
1980
1985
1990
1995
2000
2005
2010
Year
Nehalem
Dual Core Itanium 2
G6
Penryn
Itanium 2
Quad Core
Itanium 2
Xeon
Core Duo
Itanium 2
Pentium 4HT
Xeon
Dual
Pentium M
Pentium M
Pentium 4
G5
Pentium III
G4
Pentium III
Pentium MMX
G3
Pentium Pro
604e
Pentium
604
601
603e
80486
68040
80386
68030
68020
80286
68000
8086
8080
8008
6800
4004



Source: http://www.ncbi.nlm.nih.gov/Genbank/genbankstats.html

Cost per Megabase of DNA Sequence



A Database Search

go to, e.g., http://www.uniprot.org/, paste in this:

```
>sp|P15172|MYOD1_HUMAN Myoblast determination protein 1 OS=Homo
   sapiens GN=MYOD1 PE=1 SV=3
MELLSPPLRDVDLTAPDGSLCSFATTDDFYDDPCFDSPDLRFFEDLDPRLMHVGALLKPE
EHSHFPAAVHPAPGAREDEHVRAPSGHHQAGRCLLWACKACKRKTTNADRRKAATMRERR
RLSKVNEAFETLKRCTSSNPNQRLPKVEILRNAIRYIEGLQALLRDQDAAPPGAAAAFYA
PGPLPPGRGGEHYSGDSDASSPRSNCSDGMMDYSGPPSGARRRNCYEGAYYNEAPSEPRP
GKSAAVSSLDCLSSIVERISTESPAAPALLLADVPSESPPRRQEAAAPSEGESSGDPTQS
PDAAPQCPAGANPNPIYQVL
```

A Few seconds Later…

Graphical overview

Color code for identity 0-100% =

Accession	Entry name	0Query hit320	0Match hit (sqrt scale)17392	Name (Organism)
P15172	MYOD1_HUMAN			Myoblast determination protein 1 (Homo sapiens)
B2RC72	B2RC72_HUMAN			cDNA, FLJ95884, highly similar to Hom... (Homo sapiens)
E2RT59	E2RT59_CANFA			Uncharacterized protein (Canis familiaris)
P49811	MYOD1_PIG			Myoblast determination protein 1 (Sus scrofa)
D2KPI9	D2KPI9_PIG			Myogenic differentiation 1 (Sus scrofa)
F1S9A9	F1S9A9_PIG			Uncharacterized protein (Sus scrofa)
D2I0V4	D2I0V4_AILME			Putative uncharacterized protein (Ailuropoda melanoleuca)
P29331	MYOD1_SHEEP			Myoblast determination protein 1 (Ovis aries)
D2SP11	D2SP11_BUBBU			Myogenic factor MYOD1 (Bubalus bubalis)
Q0VBX9	Q0VBX9_BOVIN			Myogenic differentiation 1 (Bos taurus)
Q7YS82	MYOD1_BOVIN			Myoblast determination protein 1 (Bos taurus)
Q8C6B1	Q8C6B1_MOUSE			Myogenic differentiation 1 (Mus musculus)
A0JPK9	A0JPK9_RAT			Myogenic differentiation 1 (Rattus norvegicus)
Q02346	MYOD1_RAT			Myoblast determination protein 1 (Rattus norvegicus)
P10085	MYOD1_MOUSE			Myoblast determination protein 1 (Mus musculus)
Q6DTY5	Q6DTY5_PIG			Eukaryotic myogenic factor MYF-3 (Sus scrofa)
P21572	MYOD1_COTJA			Myoblast determination protein 1 homolog (Coturnix coturnix japonica)
Q6DV59	Q6DV59_MELGA			MyoD (Meleagris gallopavo)
P16075	MYOD1_CHICK			Myoblast determination protein 1 homolog (Gallus gallus)
C5J072	C5J072_CHICK			Myogenic differentiation 1 (Gallus gallus)
C3U0I1	C3U0I1_ANAPL			Myogenic differentiation 1 (Anas platyrhynchos)
F1NHM3	F1NHM3_CHICK			Uncharacterized protein (Gallus gallus)
F1NXM5	F1NXM5_CHICK			Uncharacterized protein (Gallus gallus)
P13904	MYODA_XENLA			Myoblast determination protein 1 homolog A (Xenopus laevis)
Q8AVZ0	Q8AVZ0_XENLA			Myod1-a protein (Xenopus laevis)
Q7T109	Q7T109_XENTR			MyoD protein (Xenopus tropicalis)

…And 975 more…

Accession	Entry name	Status	Protein names	Organism	Length
Q7T109	Q7T109_XENTR	★	**MyoD protein**	Xenopus tropicalis (Western clawed frog) (Silurana tropicalis)	288

Some Details from #25

Alignment 1 against Q7T109

Score	964	**E-value**	1.0 ×10 $^{-102}$
Identity	64.0%	**Positives**	74.0%
Query length	320	**Match length**	288
Position	**Q7T109** matches from 1 to 288 (288AA), in the query sequence from 1 to 320 (320AA)		
Graphical			

```
1     MELLSPPLRDVDLTAPDGSLCSFATTDDFYDDPCFDSPDLRFFEDLDPRLMHVGALLKPE   60 P15172
      MELL PPLRD+++T  +GSLCSF T DDFYDDPCF++ D+ FFEDLDPRL+HV ALLKPE
1     MELLPPPLRDMEVT--EGSLCSFPTPDDFYDDPCFNTSDMSFFEDLDPRLVHV-ALLKPE   57 Q7T109

61    EHSHFPAAVHPAPGAREDEHVRAPSGHHQAGRCLLWACKACKRKTTNADRRKAATMRERR  120 P15172
      +  H             EDEHVRAPSGHHQAGRCLLWACKACKRKTTNADRRKAATMRERR
58    DPHH-----------NEDEHVRAPSGHHQAGRCLLWACKACKRKTTNADRRKAATMRERR  106 Q7T109

121   RLSKVNEAFETLKRCTSSNPNQRLPKVEILRNAIRYIEGLQALLRDQDAAPPGAAAAFYA  180 P15172
      RLSKVNEAFETLKRCTS+NPNQRLPKVEILRNAIRYIE LQ+LLR Q+         +FY
107   RLSKVNEAFETLKRCTSTNPNQRLPKVEILRNAIRYIESLQSLLRGQE-------ESFY-  158 Q7T109

181   PGPLPPGRGGEHYSGDSDASSPRSNCSDGMMDYSGPPSGARRRNCYEGAYYNEAPSEPRP  240 P15172
        P+       EHYSGDSDASSPRSNCSDGM DYS PP G+RRRN Y+ ++Y+++P+  R
159   --PVL-----EHYSGDSDASSPRSNCSDGMTDYS-PPCGSRRRNSYDSSFYSDSPNGLRL  210 Q7T109

241   GKSAAVSSLDCLSSIVERISTESPAAPALLLADVPSESPPRRQEAAAPSEGES---SGDP  297 P15172
      GKS+ +SSLDCLSSIVERISTESP  P +  AD  SE  P       +P +GE+   SG
211   GKSSVISSLDCLSSIVERISTESPVCPVIPAADSGSEGSP-----CSPLQGETLSESGII  265 Q7T109
```

The foregoing search capability is a *huge* deal

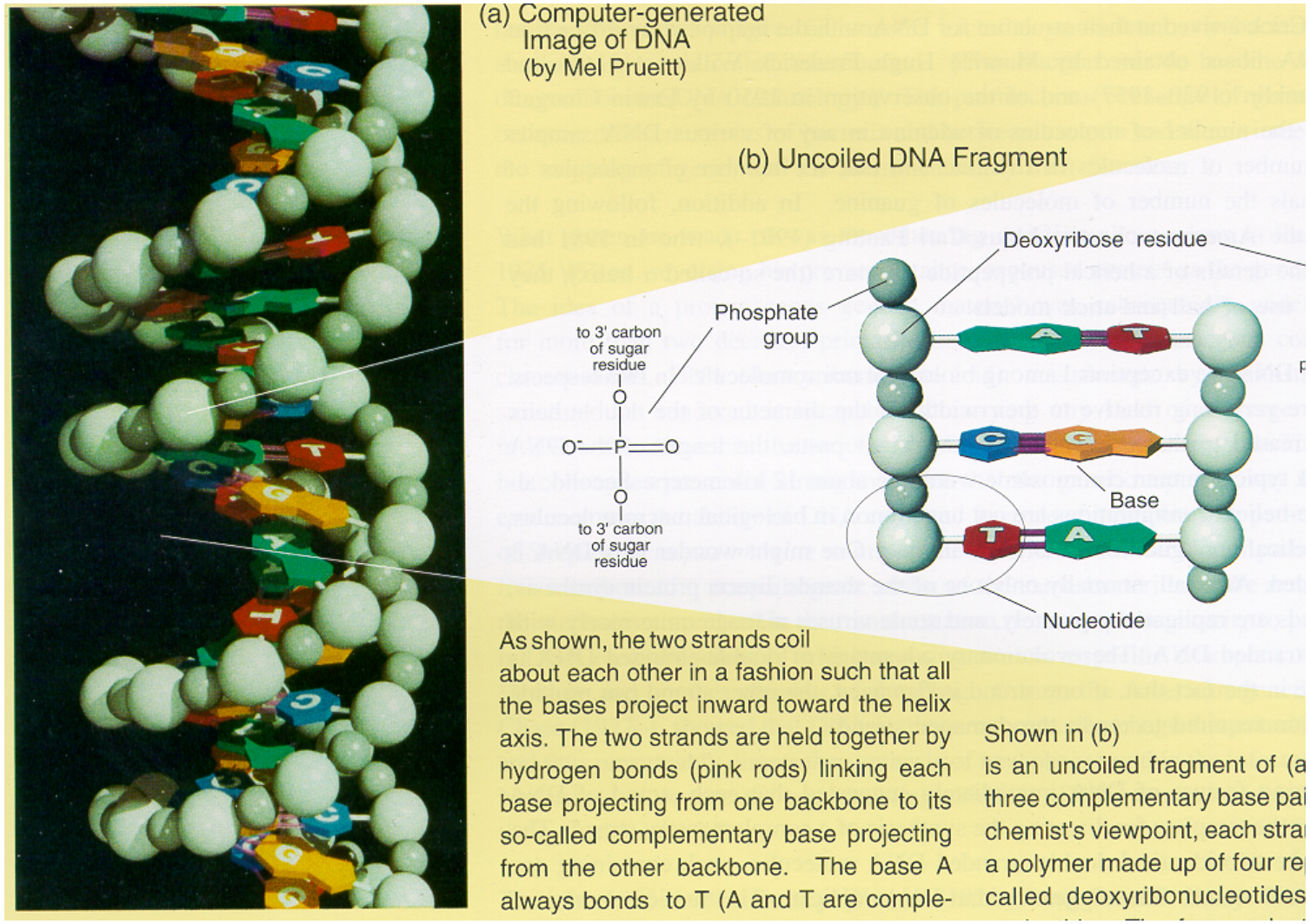
millions of searches

nearly all biologists (not just "computational biologists") use this routinely

It connects information about *all* living things

Application: RNA structure

The Double Helix



Los Alamos Science

http://www.rcsb.org/pdb/explore.do?structureId=1GAT

NATURE VOL. 227 AUGUST 8 1970

Central Dogma of Molecular Biology

by
FRANCIS CRICK
MRC Laboratory
Hills Road,
Cambridge CB2 2QH

The central dogma of molecular biology deals with the detailed residue-by-residue transfer of sequential information. It states that such information cannot be transferred from protein to either protein or nucleic acid.

"The central dogma, enunciated by Crick in 1958 and the keystone of molecular biology ever since, is likely to prove a considerable over-simplification."

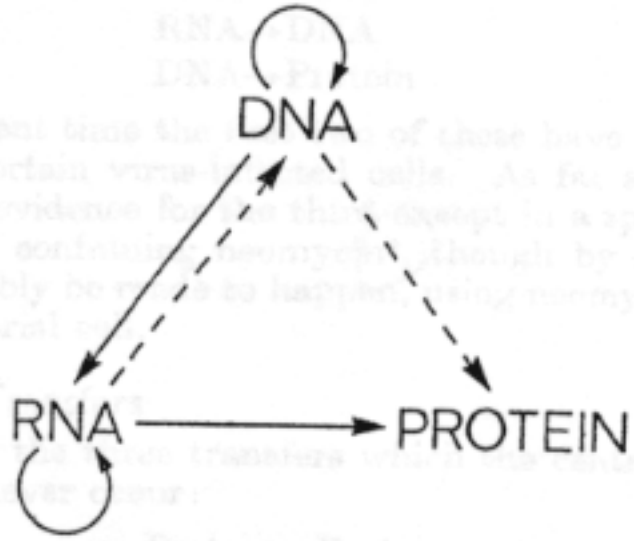


Fig. 2. The arrows show the situation as it seemed in 1958. Solid arrows represent probable transfers, dotted arrows possible transfers. The absent arrows (compare Fig. 1) represent the impossible transfers postulated by the central dogma. They are the three possible arrows starting from protein.

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)

Since mid 90's dramatic discoveries

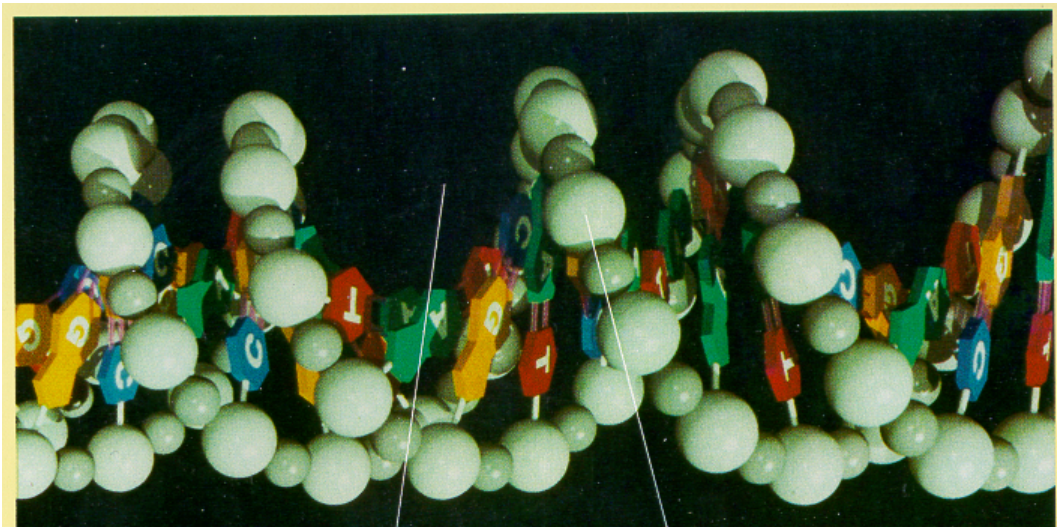
- Regulation, transport, stability/degradation
- E.g. "microRNA": 100s in humans => 50% of genes
- E.g. "riboswitches": 1000s in bacteria

DNA structure: dull

5’ …ACCGCTAGATG…3’

| | | | | | | | | | |

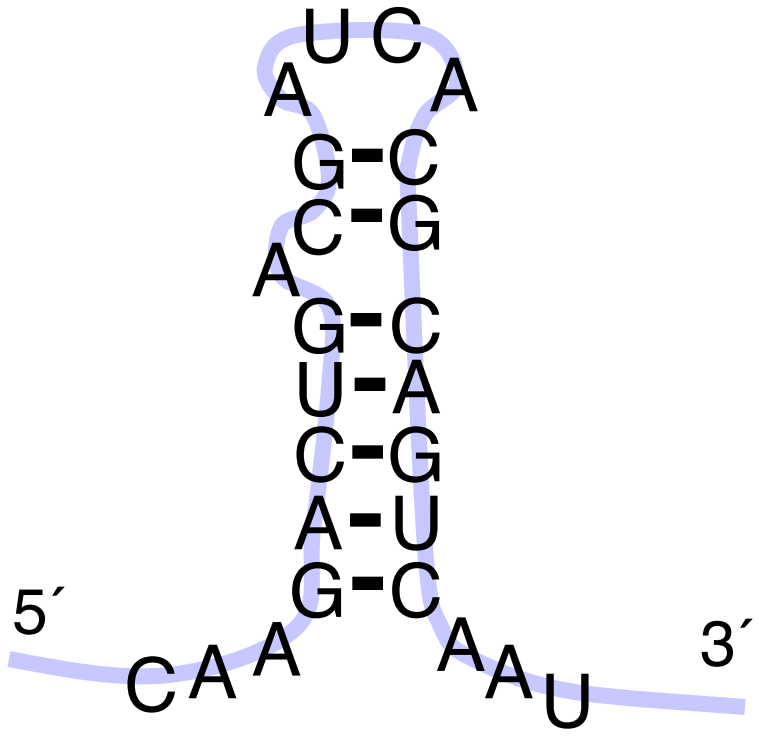
3’ …TGGCGATCTAC…5’



RNA Secondary Structure:
RNA makes helices too

Base pairs

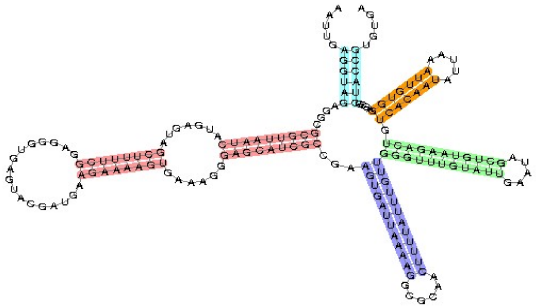
A-U
C-G

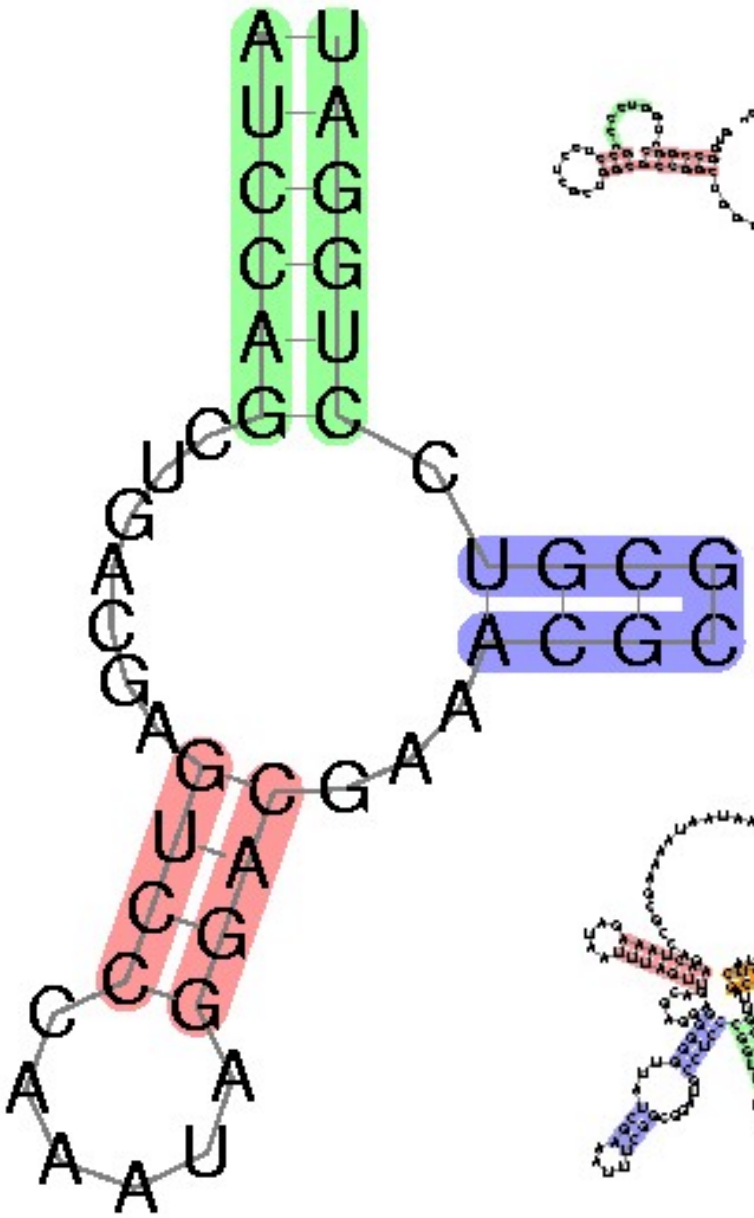


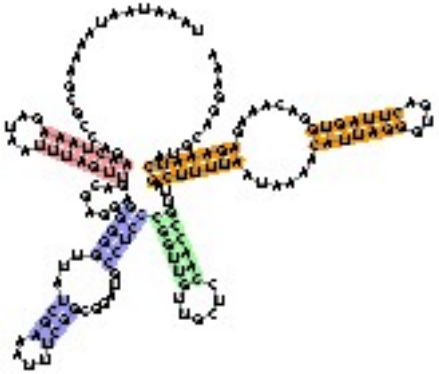
Usually *single* stranded

RNA Secondary Structure:

Not everything, but important, easier than 3d

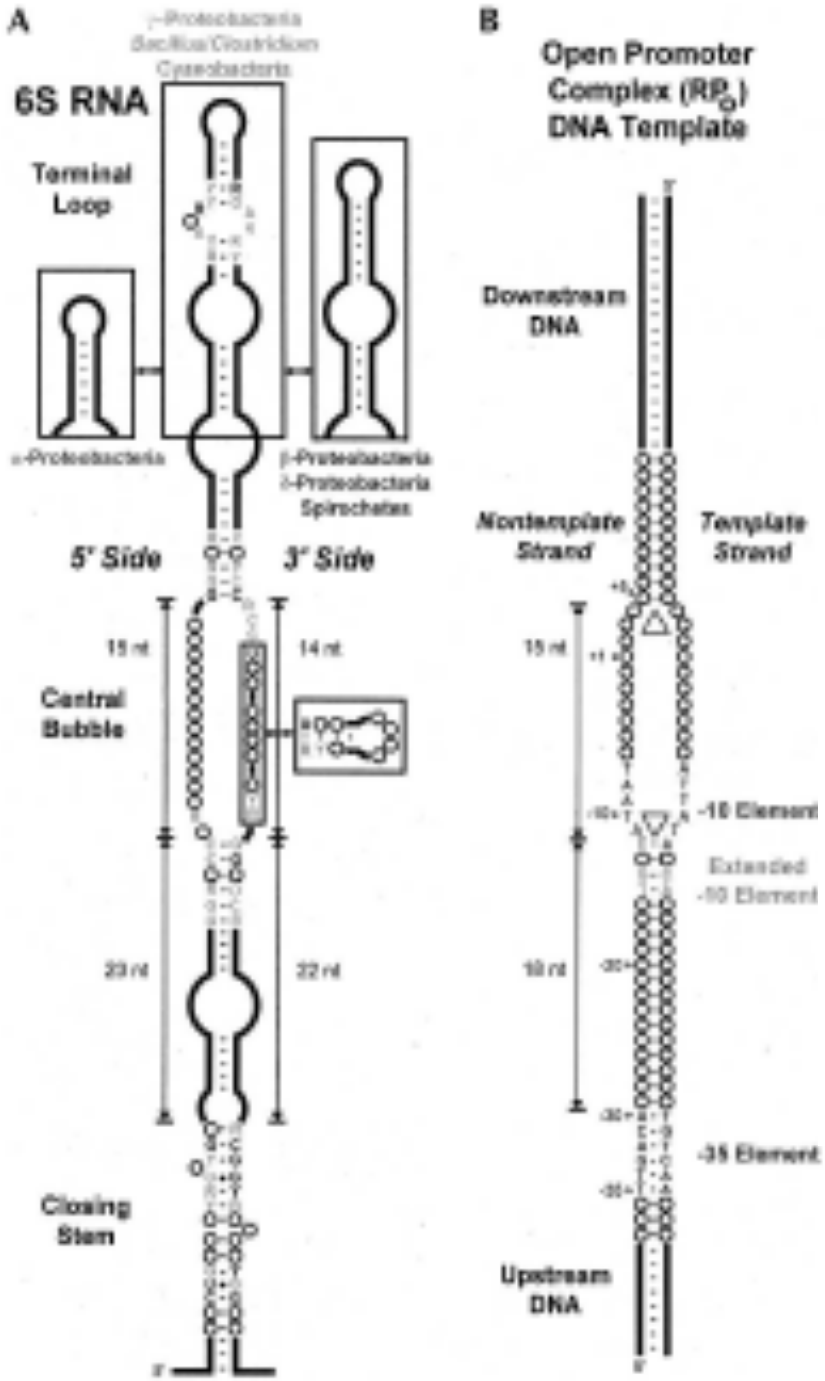




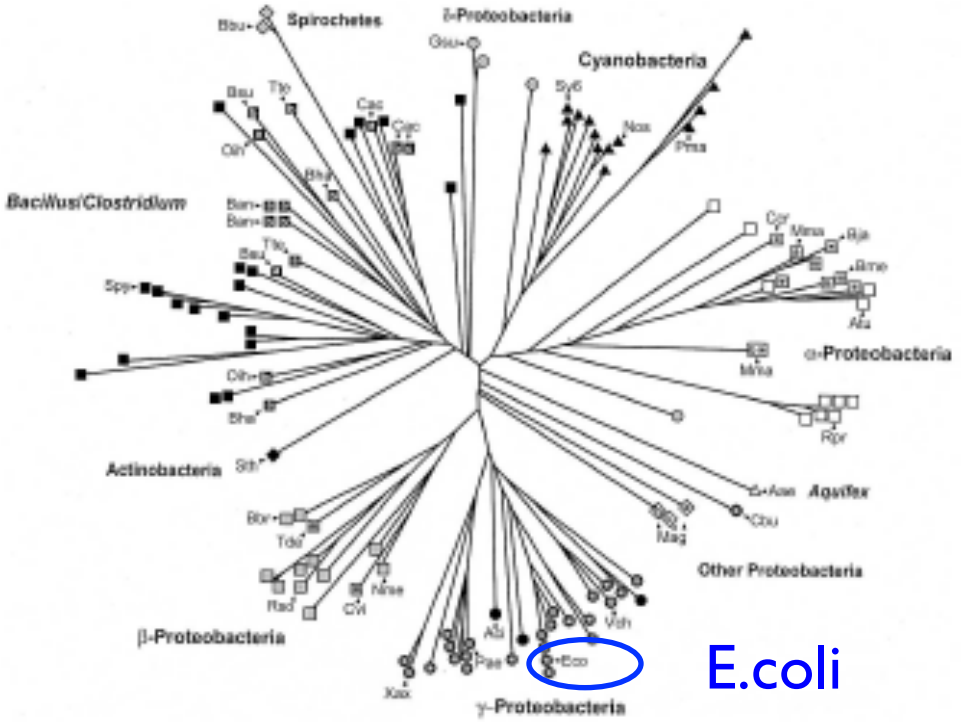


Why is structure important?

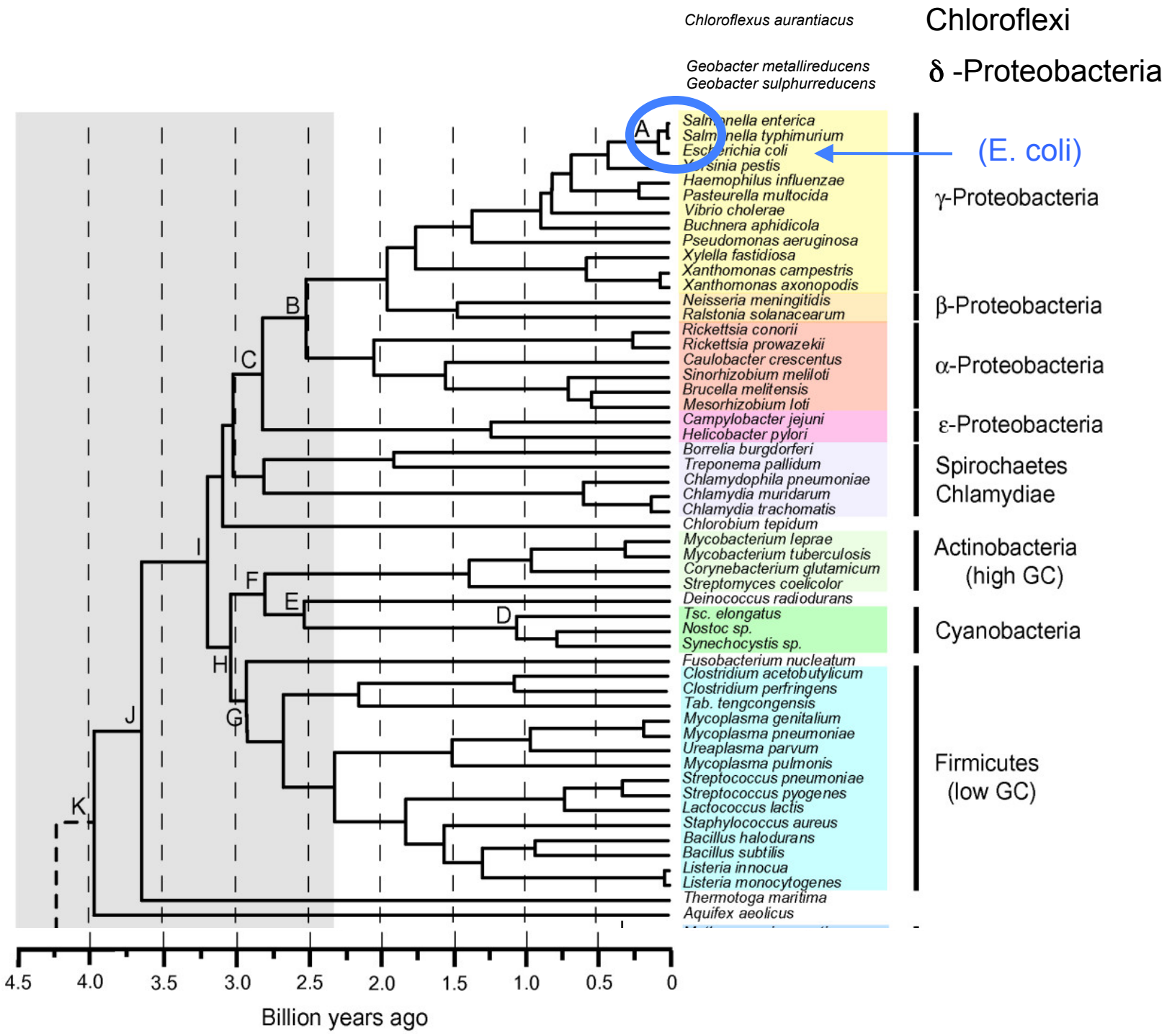
- For protein-coding, similarity in sequence is a powerful tool for finding related sequences
  - e.g. "hemoglobin" is easily recognized in all vertebrates
- For non-coding RNA, many different sequences have the same structure, and structure is most important for function.
  - So, using structure plus sequence, can find related sequences at much greater evolutionary distances



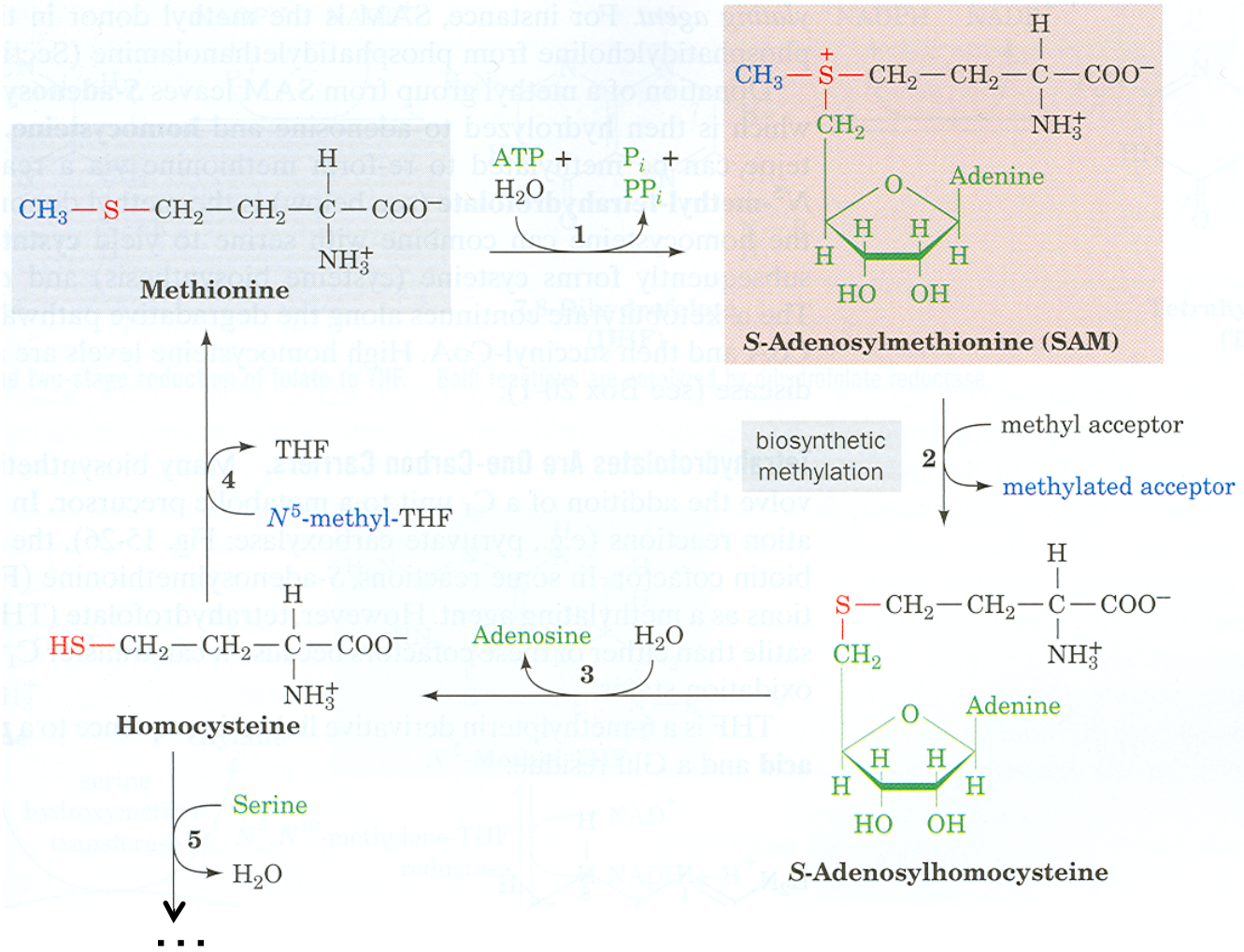
6S mimics an open promoter



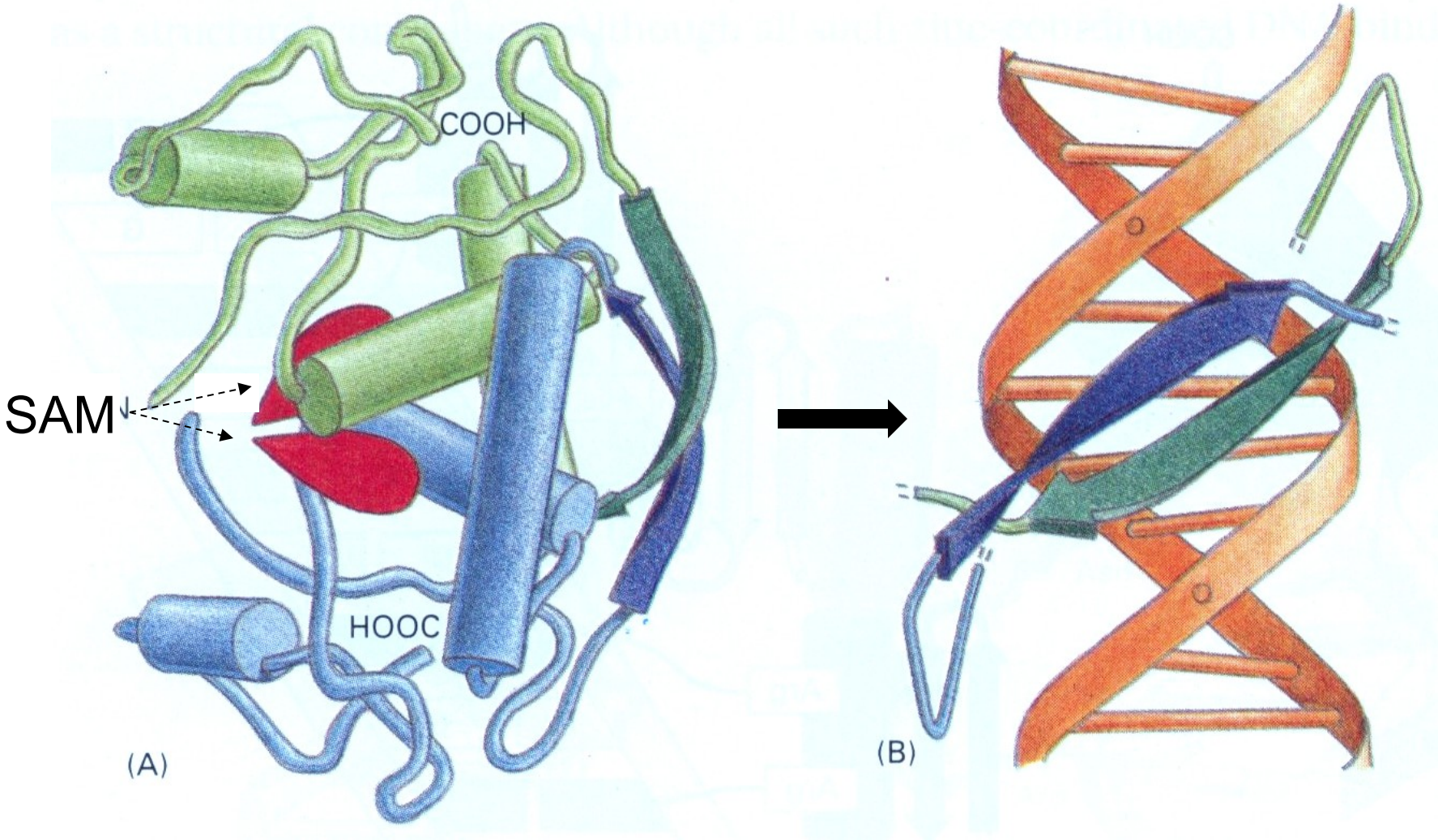
Barrick et al. *RNA* 2005
Trotochaud et al. *NSMB* 2005
Willkomm et al. NAR 2005


Chloroflexus aurantiacus
Chloroflexi
Geobacter metallireducens
Geobacter sulphurreducens
δ -Proteobacteria
A
B
C
D
E
F
G
H
I
J
K
Salmonella enterica
Salmonella typhimurium
Escherichia coli
Yersinia pestis
Haemophilus influenzae
Pasteurella multocida
Vibrio cholerae
Buchnera aphidicola
Pseudomonas aeruginosa
Xylella fastidiosa
Xanthomonas campestris
Xanthomonas axonopodis
Neisseria meningitidis
Ralstonia solanacearum
Rickettsia conorii
Rickettsia prowazekii
Caulobacter crescentus
Sinorhizobium meliloti
Brucella melitensis
Mesorhizobium loti
Campylobacter jejuni
Helicobacter pylori
Borrelia burgdorferi
Treponema pallidum
Chlamydophila pneumoniae
Chlamydia muridarum
Chlamydia trachomatis
Chlorobium tepidum
Mycobacterium leprae
Mycobacterium tuberculosis
Corynebacterium glutamicum
Streptomyces coelicolor
Deinococcus radiodurans
Tsc. elongatus
Nostoc sp.
Synechocystis sp.
Fusobacterium nucleatum
Clostridium acetobutylicum
Clostridium perfringens
Tab. tengcongensis
Mycoplasma genitalium
Mycoplasma pneumoniae
Ureaplasma parvum
Mycoplasma pulmonis
Streptococcus pneumoniae
Streptococcus pyogenes
Lactococcus lactis
Staphylococcus aureus
Bacillus halodurans
Bacillus subtilis
Listeria innocua
Listeria monocytogenes
Thermotoga maritima
Aquifex aeolicus
(E. coli)
γ-Proteobacteria
β-Proteobacteria
α-Proteobacteria
ε-Proteobacteria
Spirochaetes
Chlamydiae
Actinobacteria
(high GC)
Cyanobacteria
Firmicutes
(low GC)
4.5
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0
Billion years ago

In Bacteria: A typical biosynthetic cycle around a critical metabolite ("SAM")



Gene Regulation: The MET Repressor

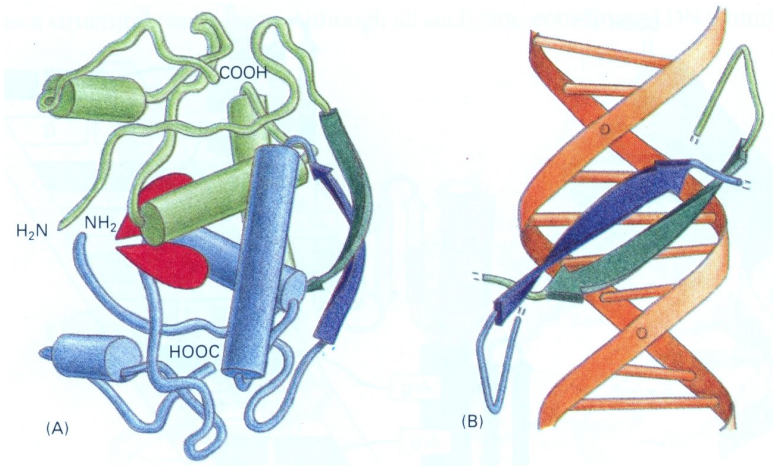


Protein

Alberts, et al, 3e.

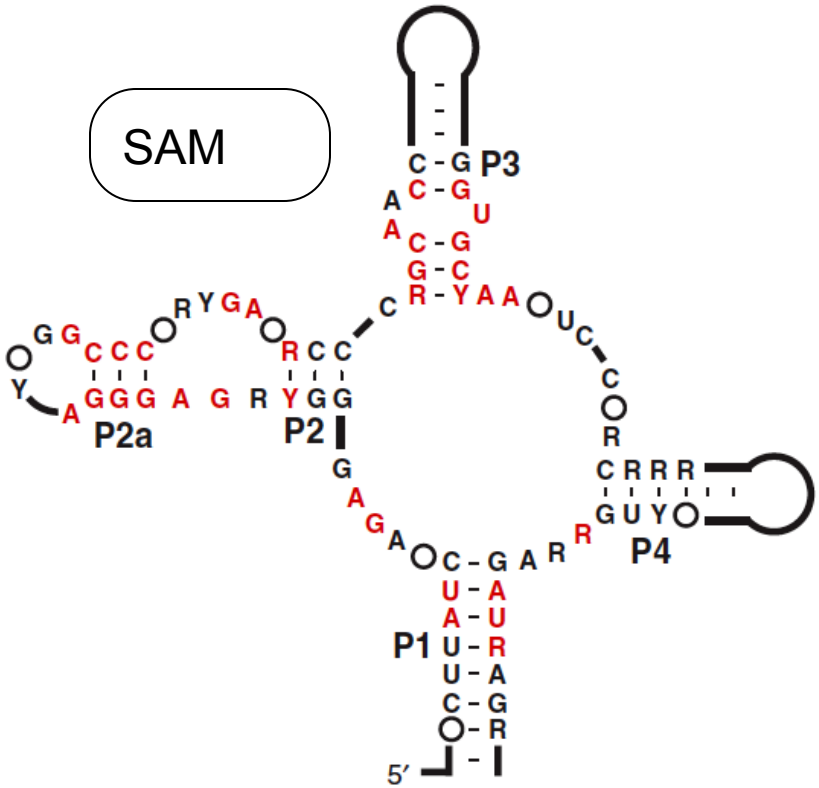
DNA

Alberts, et al, 3e.



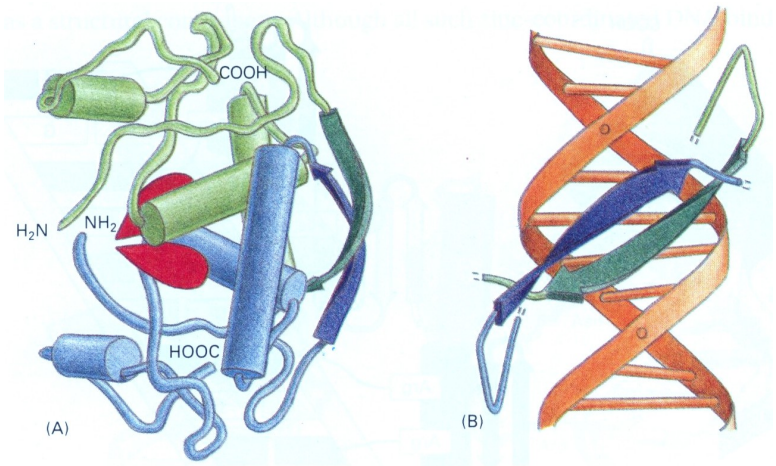
The protein way

Riboswitch alternative



Grundy & Henkin, Mol. Microbiol 1998
Epshtein, et al., PNAS 2003
Winkler et al., Nat. Struct. Biol. 2003

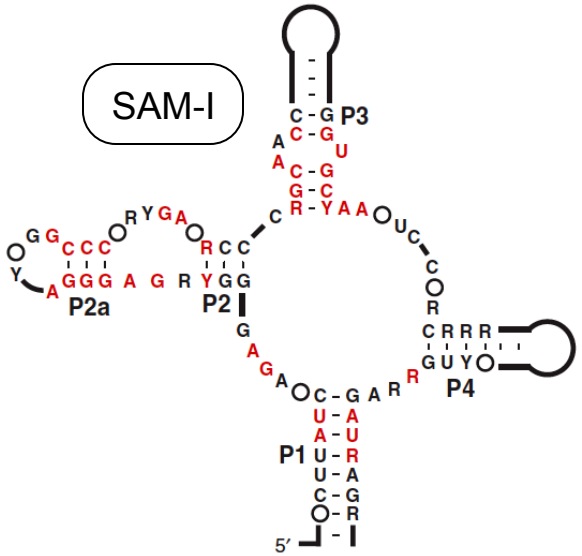
Alberts, et al, 3e.



The protein way

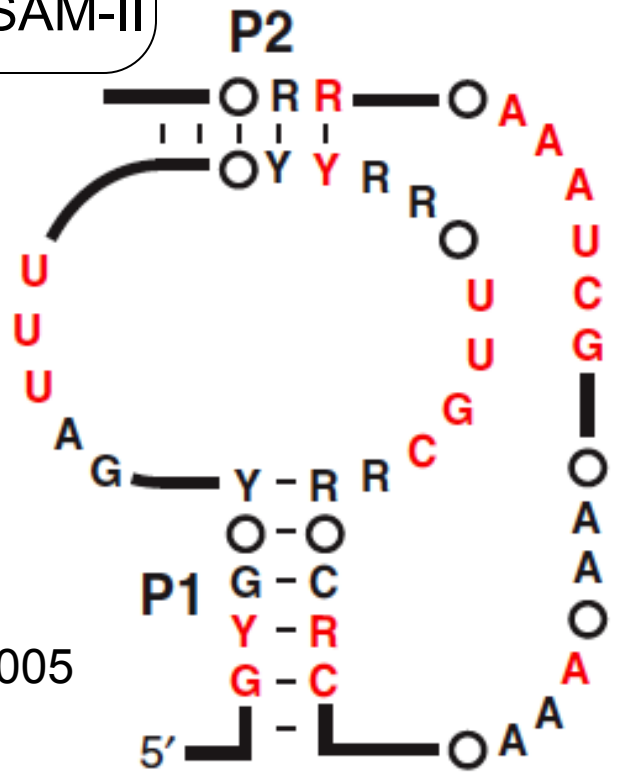
Riboswitch alternatives

SAM-I



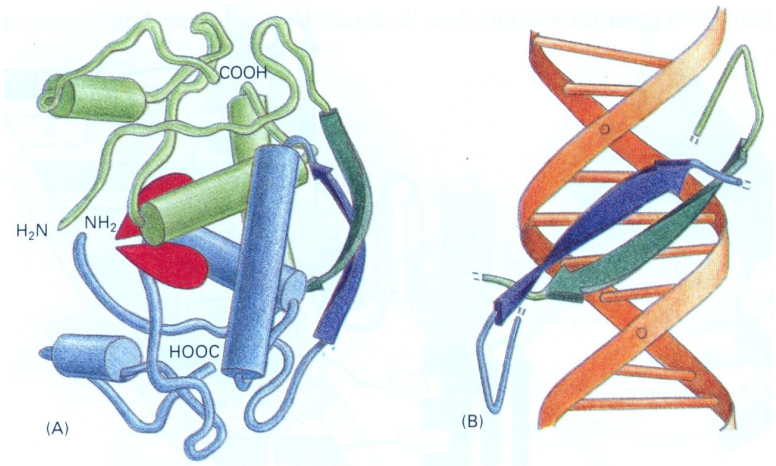
Grundy, Epshtein, Winkler et al., 1998, 2003

SAM-II



Corbino et al., Genome Biol. 2005

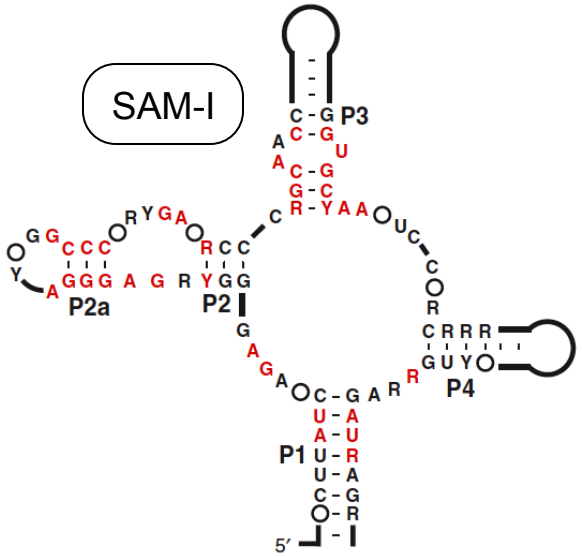
Alberts, et al, 3e.



The protein way

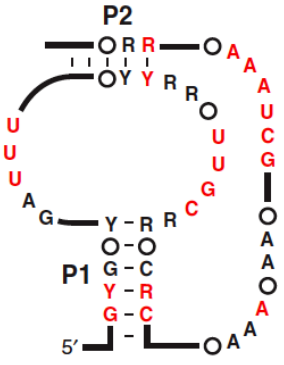
Riboswitch alternatives

SAM-I



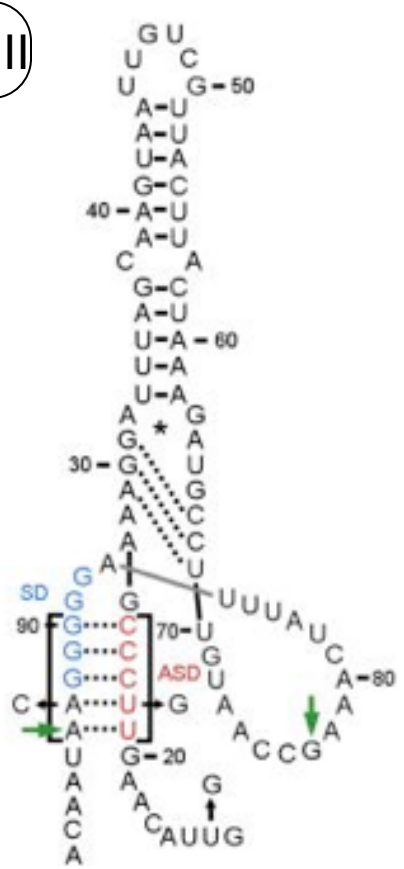
Grundy, Epshtein, Winkler et al., 1998, 2003

SAM-II



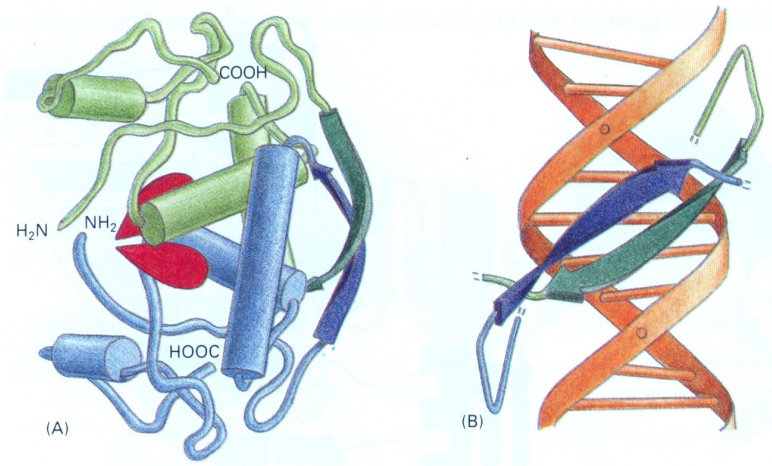
Corbino et al., Genome Biol. 2005

SAM-III



Fuchs et al., NSMB 2006

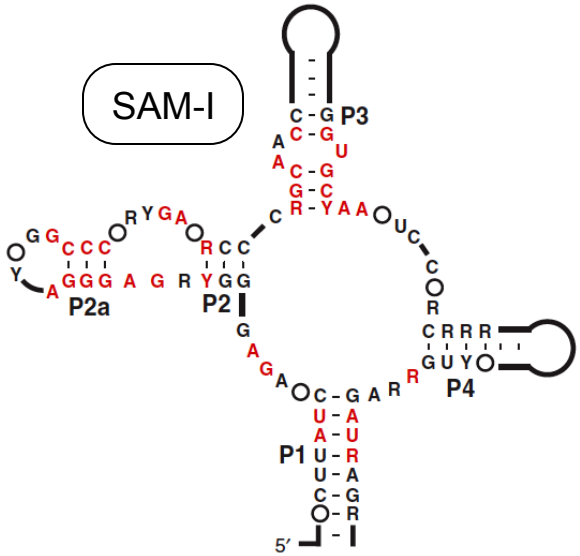
Alberts, et al, 3e.



The protein way

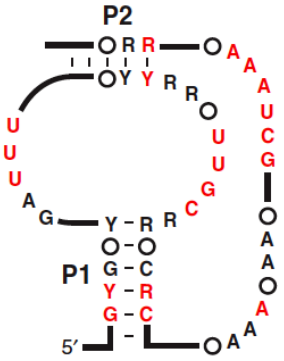
Riboswitch alternatives

SAM-I



Grundy, Epshtein, Winkler et al., 1998, 2003

SAM-II

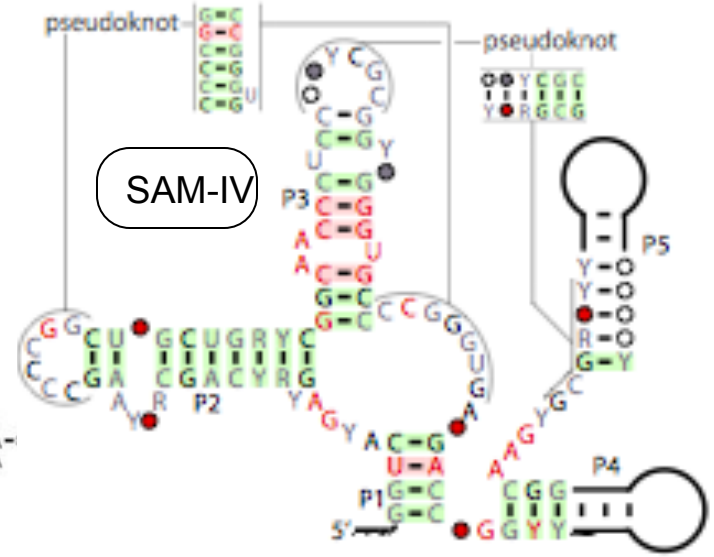


Corbino et al., Genome Biol. 2005

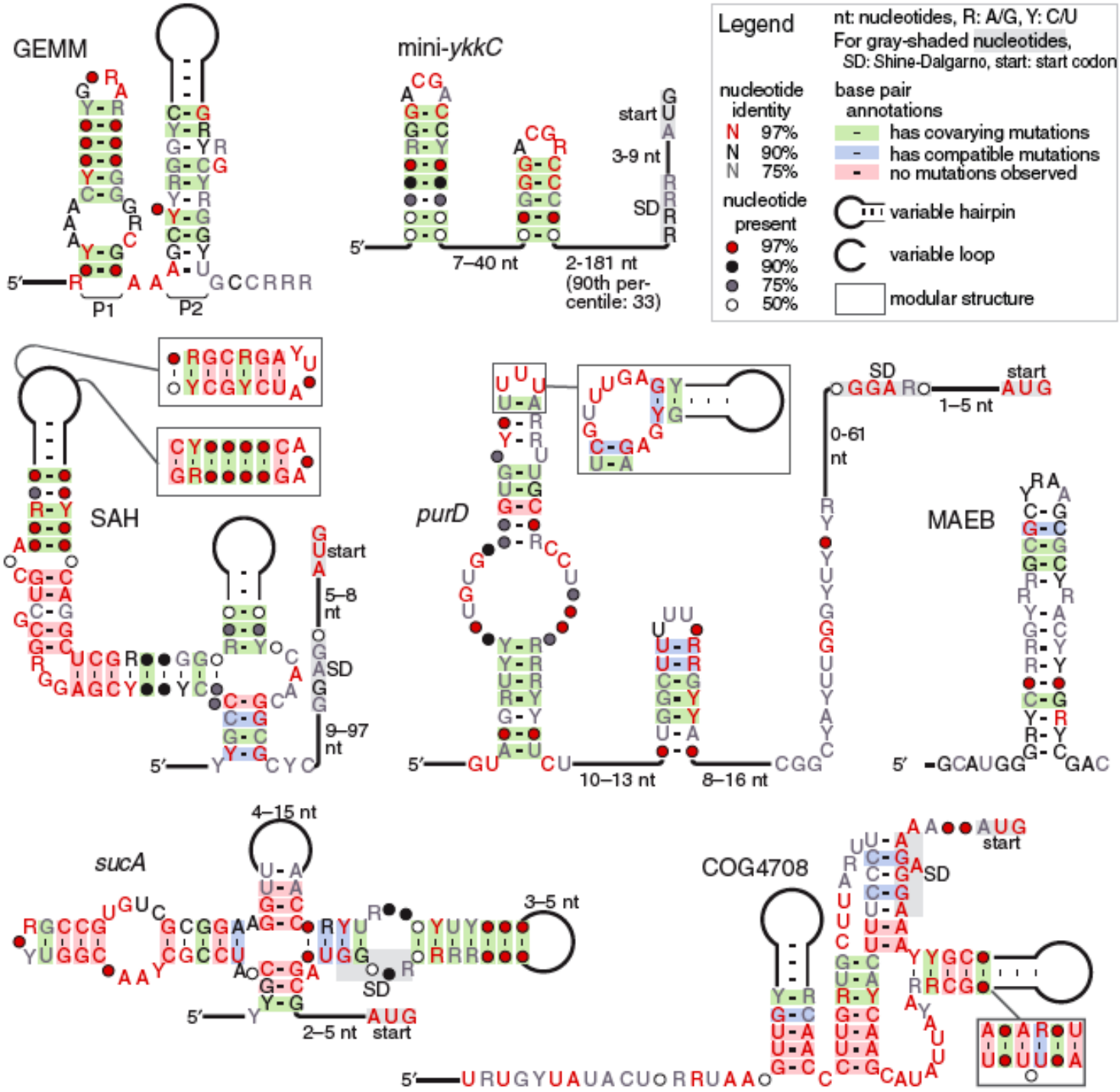
SAM-III

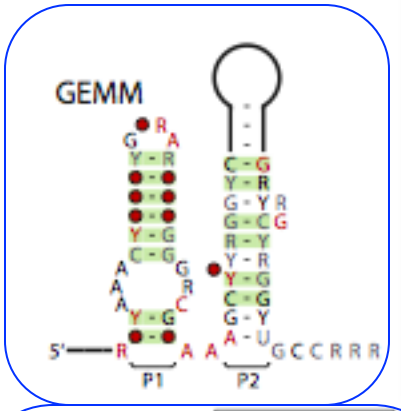
Fuchs et al., NSMB 2006

SAM-IV

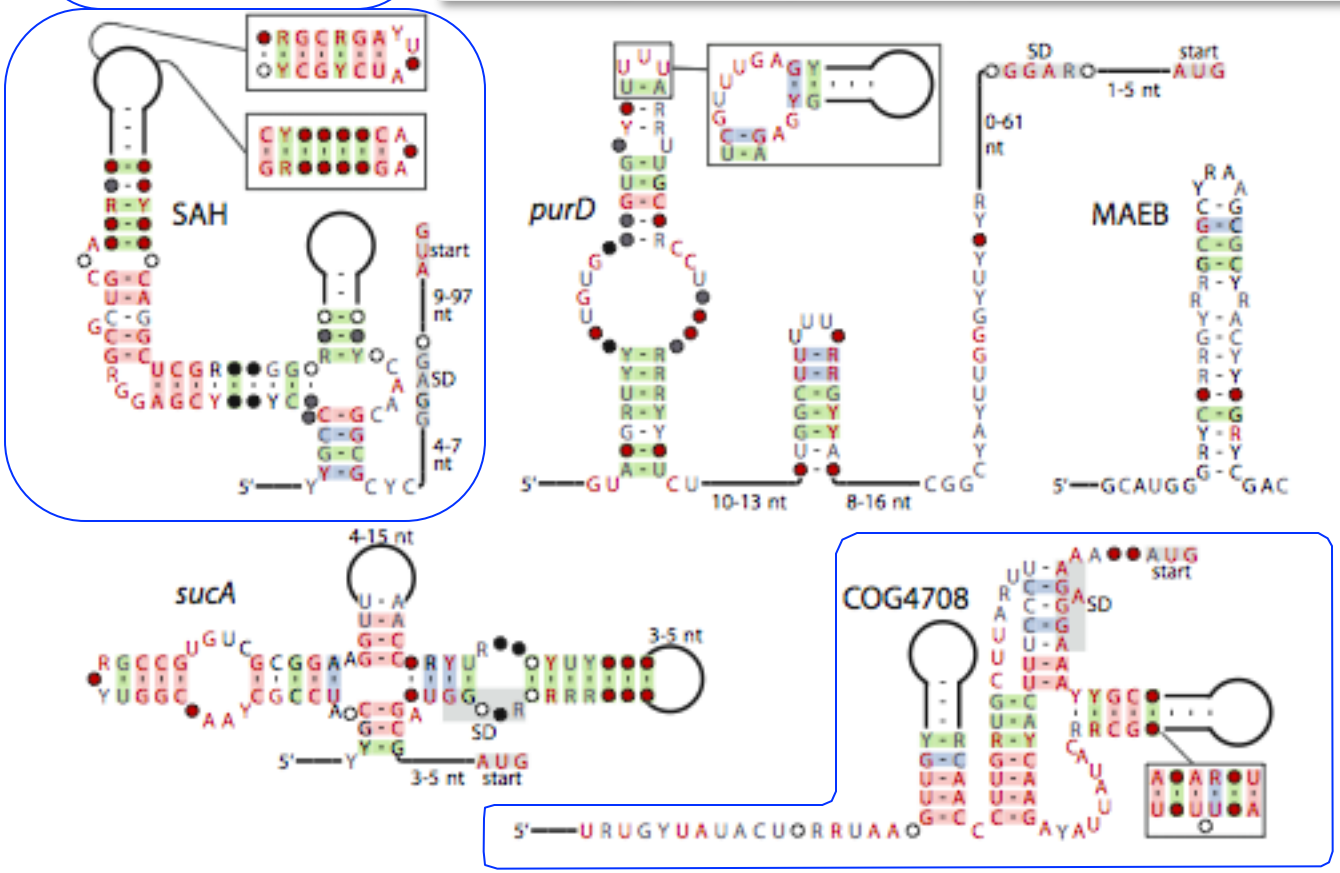


Weinberg et al., RNA 2008


GEMM
P1
P2
mini-ykkC
start
3-9 nt
SD
7–40 nt
2-181 nt
(90th per-
centile: 33)
Legend
nt: nucleotides, R: A/G, Y: C/U
For gray-shaded nucleotides,
SD: Shine-Dalgarno, start: start codon
nucleotide identity
N 97%
N 90%
N 75%
nucleotide present
97%
90%
75%
50%
base pair annotations
has covarying mutations
has compatible mutations
no mutations observed
variable hairpin
variable loop
modular structure
SAH
5–8 nt
9–97 nt
purD
10–13 nt
8–16 nt
SD
1–5 nt
0-61 nt
MAEB
sucA
4–15 nt
3–5 nt
2–5 nt
COG4708



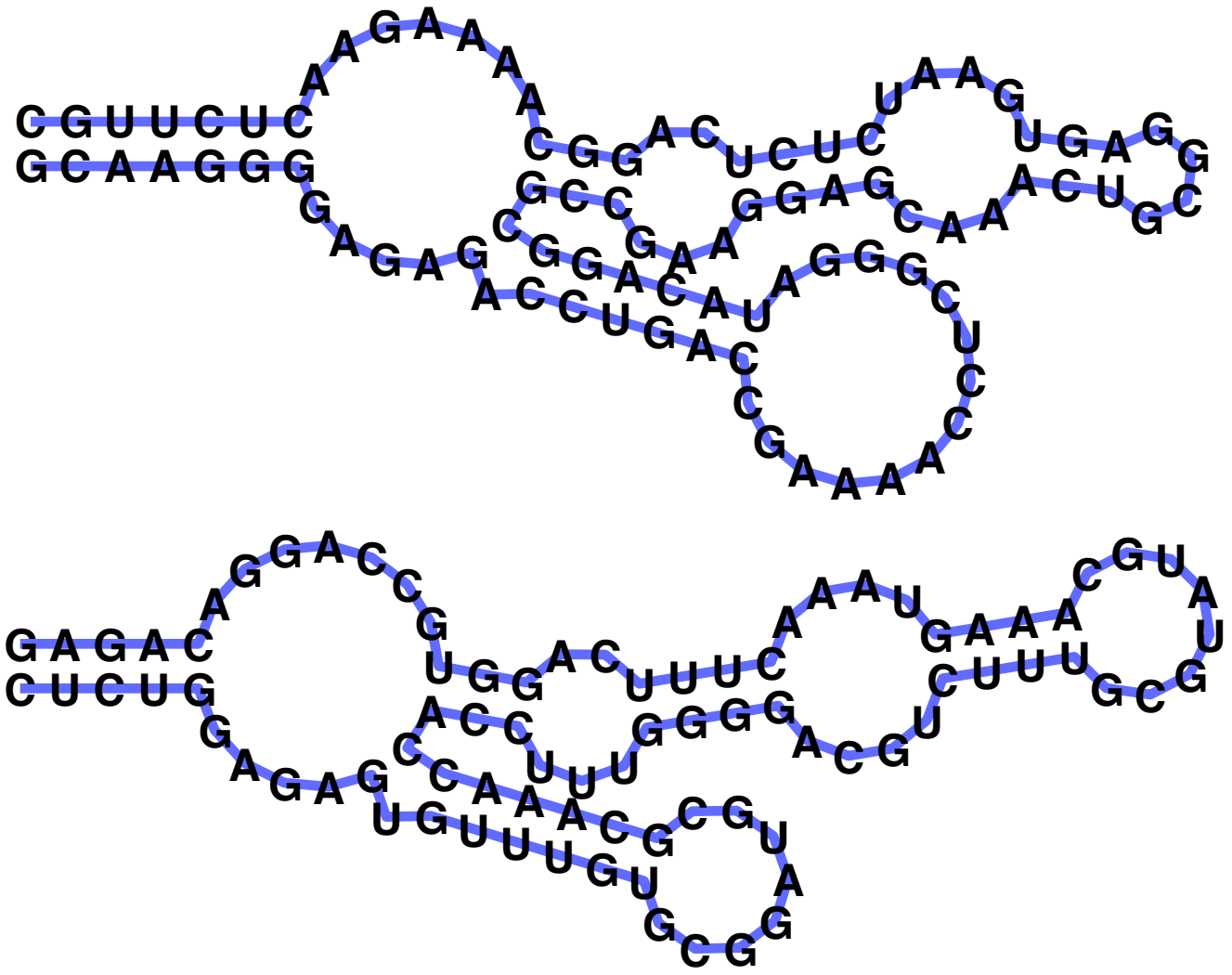
And many other examples. Widespread, deeply conserved, structurally sophisticated, functionally diverse, biologically important uses for ncRNA throughout prokaryotic world.



boxed = confirmed riboswitch (+2 more)

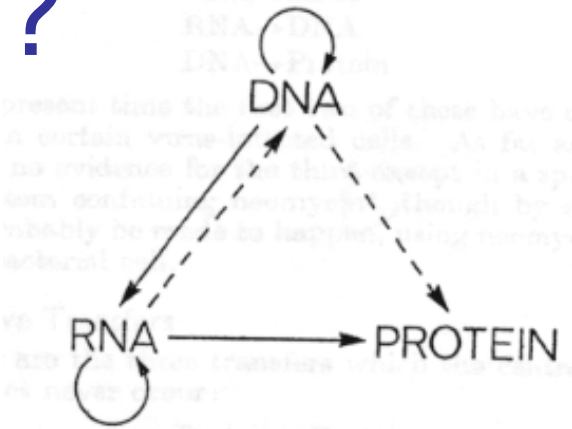
Weinberg, et al. Nucl. Acids Res., July 2007 35: 4809-4819.

Q: What's so hard?



A: Structure often more important than sequence

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

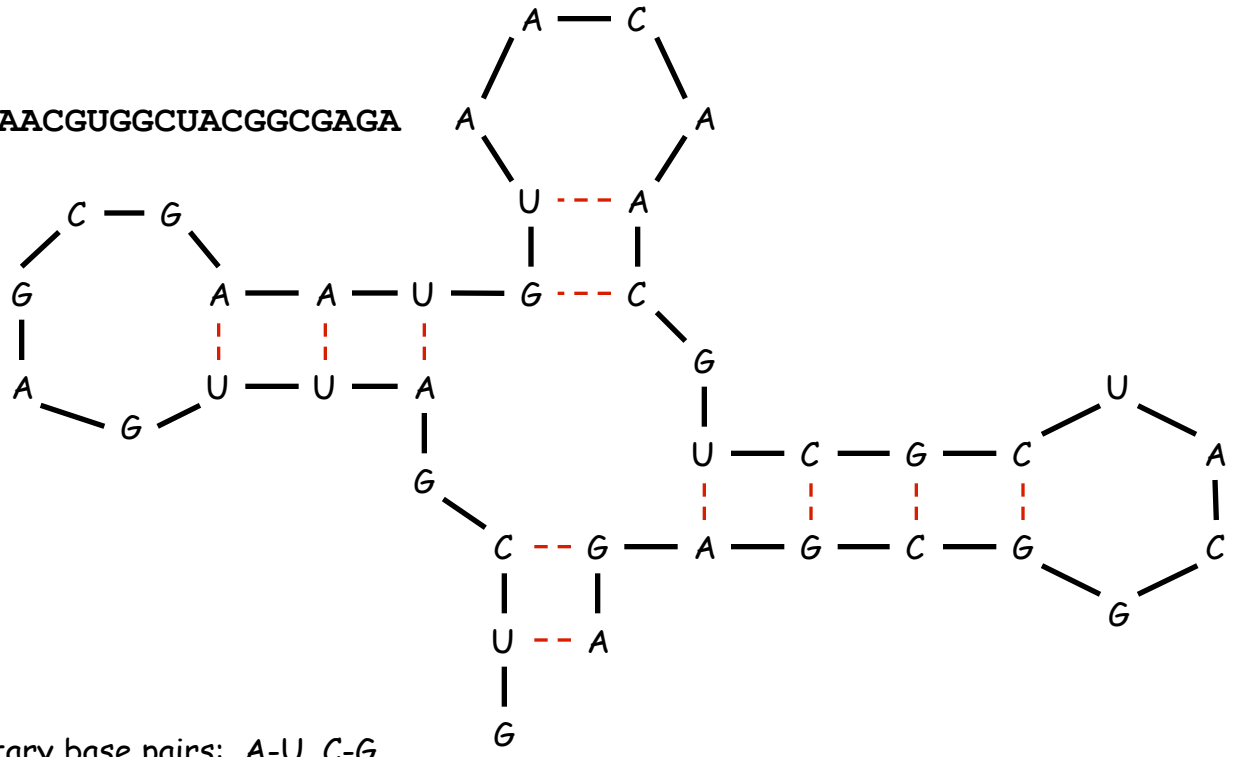
Nussinov's Algorithm – core technology
for RNA structure prediction

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 behavior of molecule.

Ex: **GUCGAUUGAGCGAAUGUAACAACGUGGCUACGGCGAGA**



complementary base pairs: A-U, C-G

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.*)

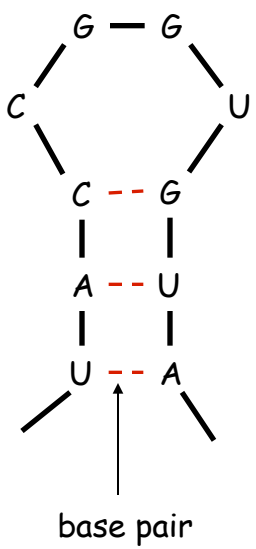
Free energy. Usual hypothesis is that an RNA molecule will form the secondary structure with the optimum total free energy.

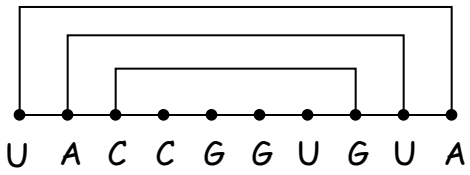
approximate by number of base pairs

Goal. Given an RNA molecule $B = b_1 b_2 \ldots b_n$, find a secondary structure S that maximizes the number of base pairs.

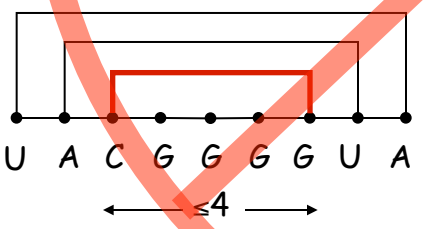
RNA Secondary Structure: Examples

Examples.

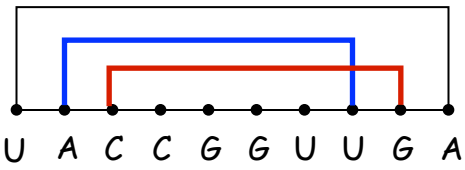


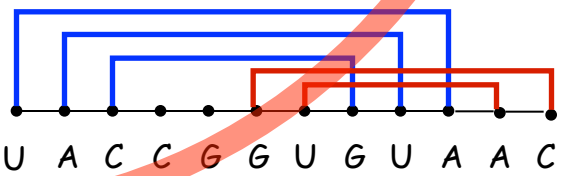


ok



sharp turn

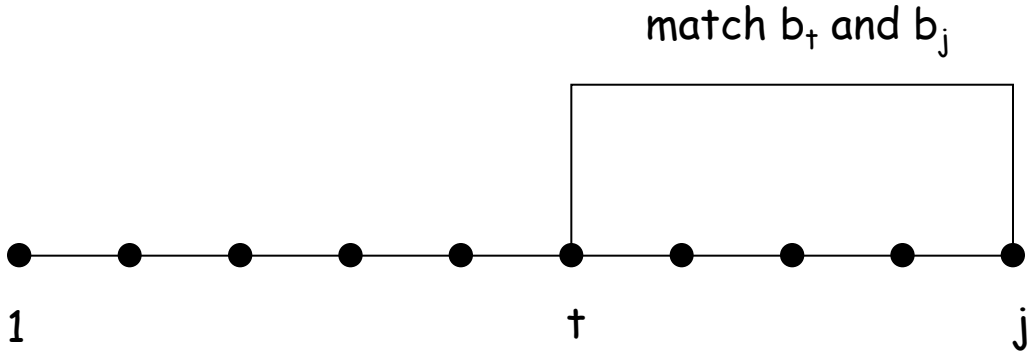




crossing

RNA Secondary Structure: Subproblems

First attempt. OPT[j] = maximum number of base pairs in a secondary structure of the substring $b_1 b_2 \ldots b_j$.



Difficulty. Results in two sub-problems.

- Finding secondary structure in: $b_1 b_2 \ldots b_{t-1}$. ← OPT(t-1)
- Finding secondary structure in: $b_{t+1} b_{t+2} \ldots b_{j-1}$. ← not OPT of anything; need more sub-problems

Dynamic Programming Over Intervals: (R. Nussinov's algorithm)

Notation. OPT[i, j] = maximum number of base pairs in a secondary structure of the substring $b_i b_{i+1} \ldots b_j$.

- Case 1. If $i \geq j - 4$.

  OPT[i, j] = 0 by no-sharp turns condition.

- Case 2. Base b_j is not involved in a pair.

  OPT[i, j] = 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

$$OPT[i, j] = 1 + \max_t \{ OPT[i, t-1] + OPT[t+1, j-1] \}$$

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)

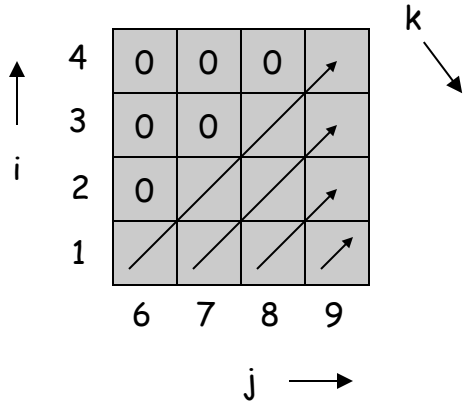
Remark. Same core idea in CKY algorithm to parse context-free grammars.

Bottom Up Dynamic Programming Over Intervals

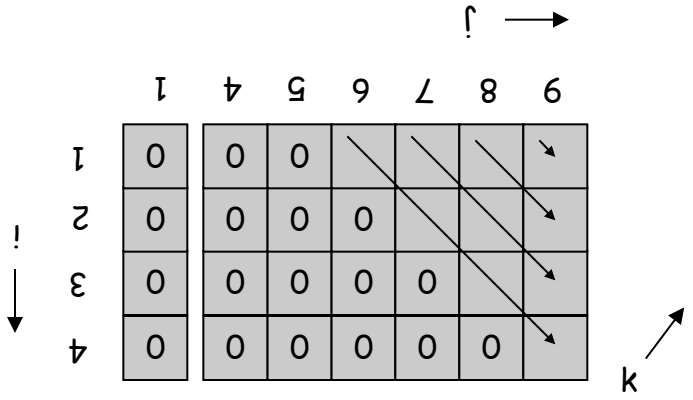
Q. What order to solve the sub-problems?

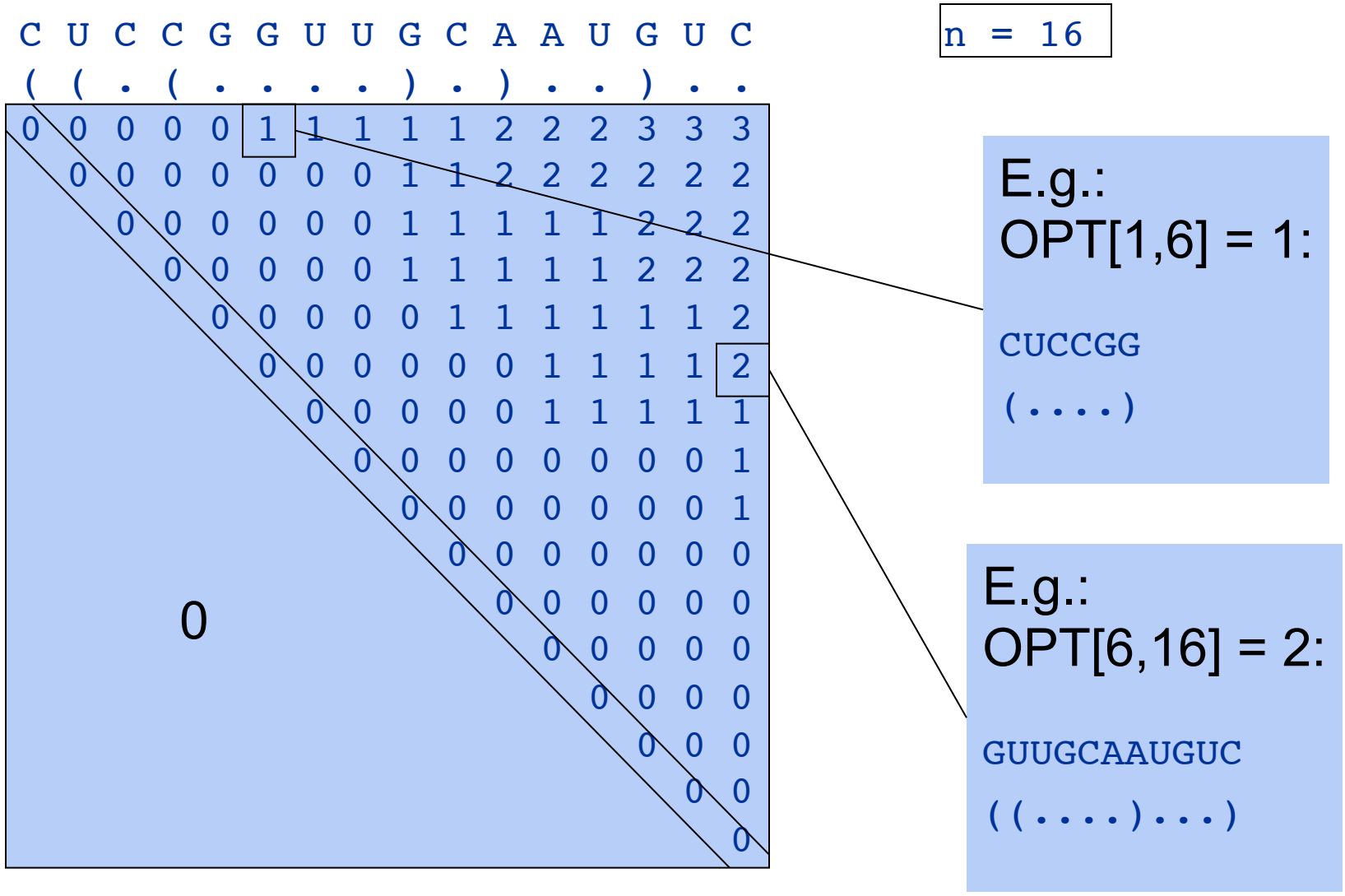
A. Do shortest intervals first.

```
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]
}
```

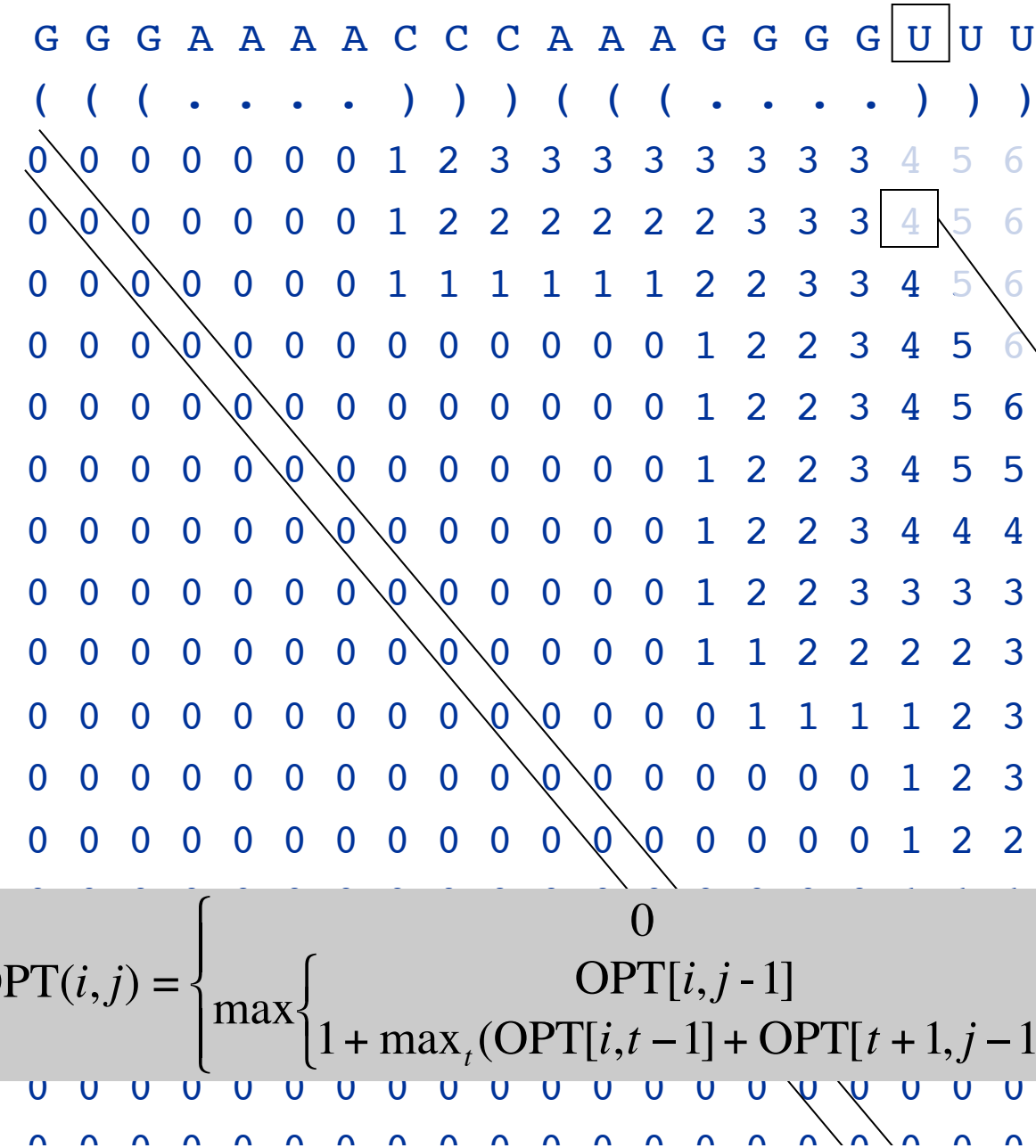


Running time. $O(n^3)$.




C U C C G G U U G C A A U G U C
((. (. . . .) .) . .) . .
0 0 0 0 0 1 1 1 1 1 2 2 2 3 3 3
0 0 0 0 0 0 0 1 1 2 2 2 2 2 2
0 0 0 0 0 0 1 1 1 1 1 2 2 2
0 0 0 0 0 1 1 1 1 1 2 2 2
0 0 0 0 0 1 1 1 1 1 1 2
0 0 0 0 0 0 1 1 1 1 2
0 0 0 0 0 1 1 1 1 1
0 0 0 0 0 0 0 0 1
0 0 0 0 0 0 0 0 1
0 0 0 0 0 0 0 0
0 0 0 0 0 0 0
0 0 0 0 0 0
0 0 0 0 0
0 0 0 0
0 0 0
0 0
0
0
n = 16
E.g.:
OPT[1,6] = 1:
CUCCGG
(....)
E.g.:
OPT[6,16] = 2:
GUUGCAAUGUC
((....)...)

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



Case 1:
2 ≥ 18-4? no.

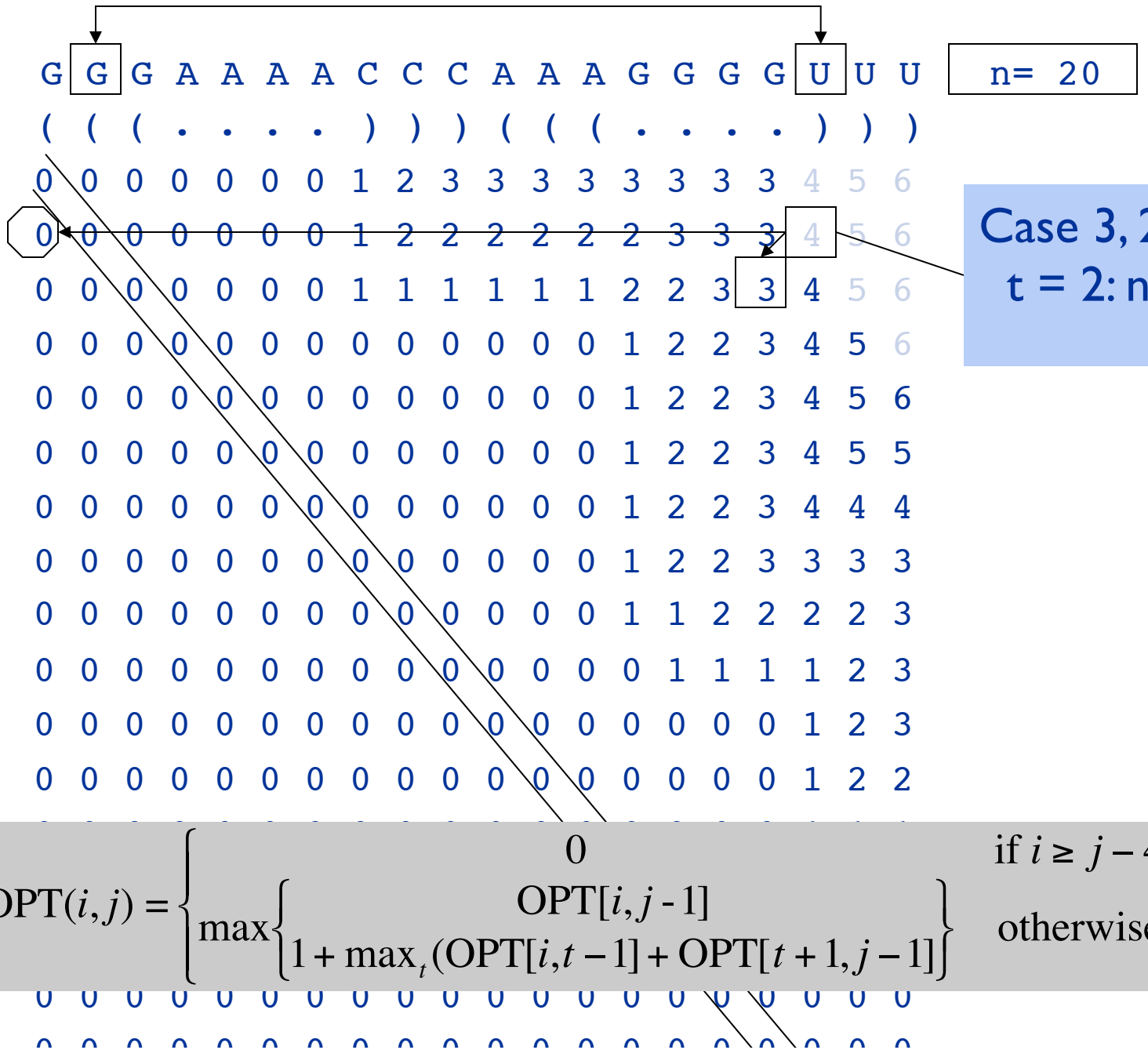
Case 2:
B_{18} unpaired?
Always a possibility;
then OPT[2,18] ≥ 3

GGAAAACCCAAAGGGGU

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

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

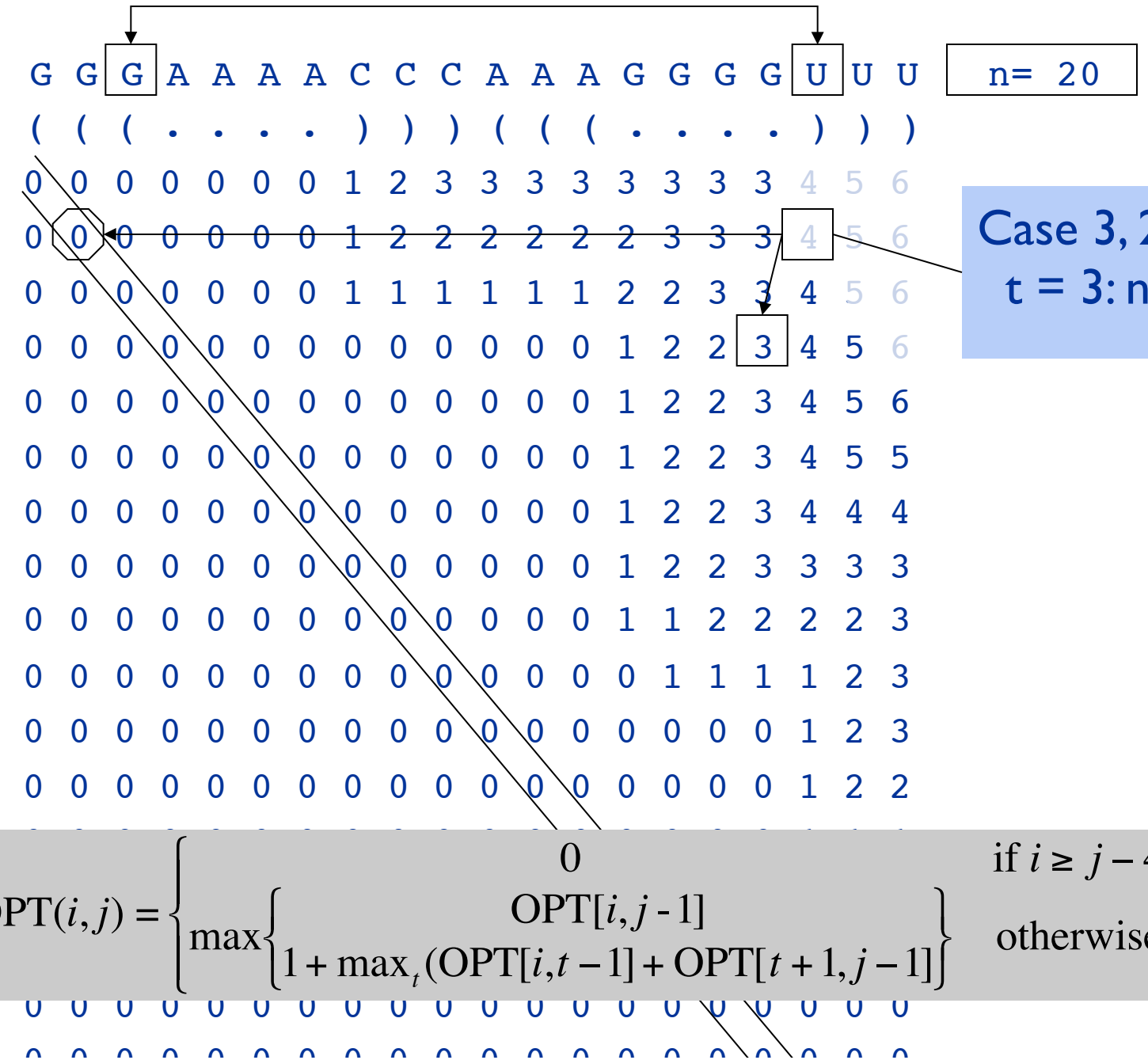
Computing one cell: OPT[2,18] = ?



Case 3, $2 \le t < 18-4$:
t = 2: no pair

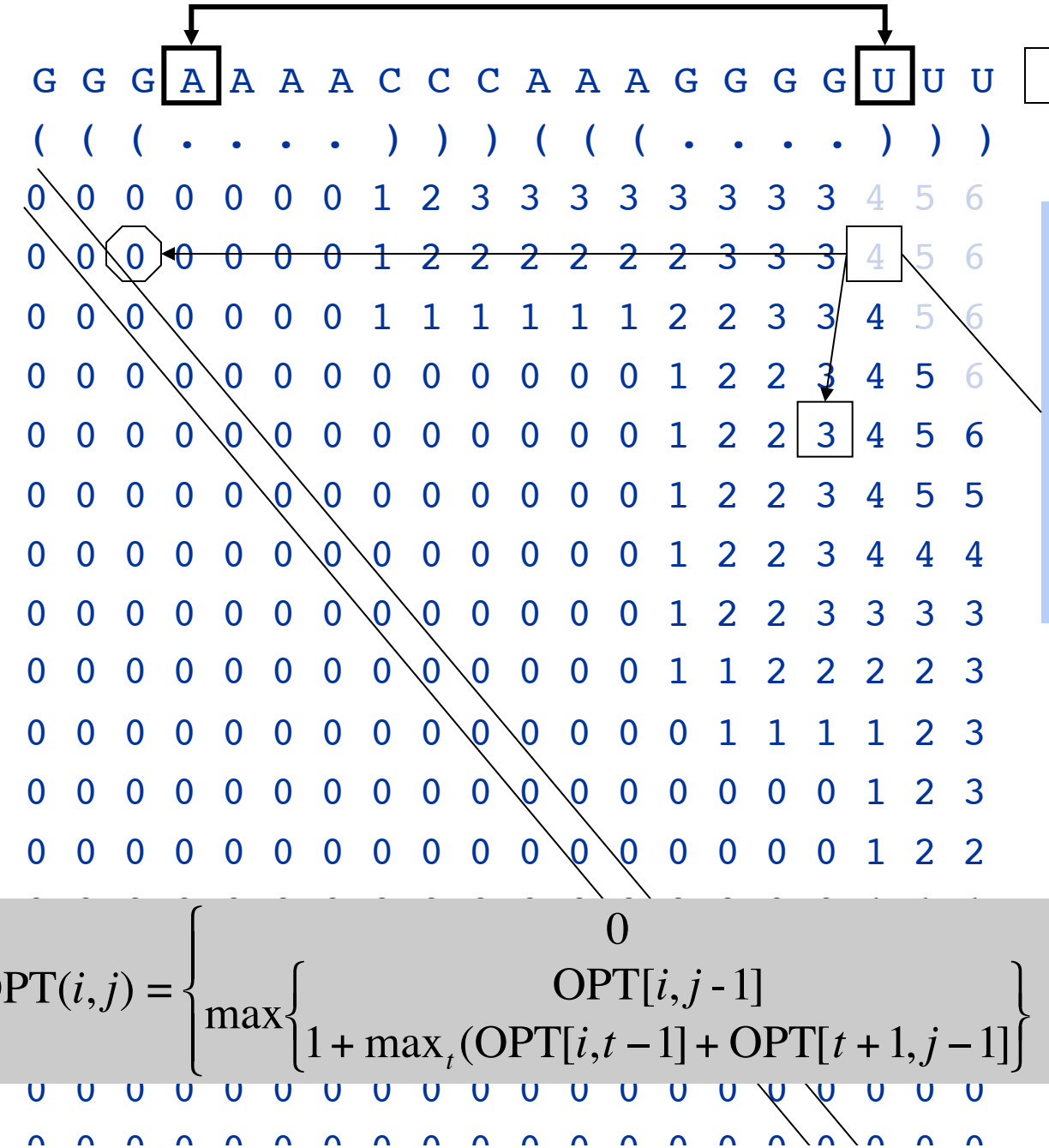
$$\mathrm{OPT}(i,j) = \begin{cases} 0 & \text{if } i \ge j-4 \\ \max\left\{ \begin{array}{c} \mathrm{OPT}[i,j-1] \\ 1+\max_t(\mathrm{OPT}[i,t-1]+\mathrm{OPT}[t+1,j-1]) \end{array} \right\} & \text{otherwise} \end{cases}$$

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



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

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

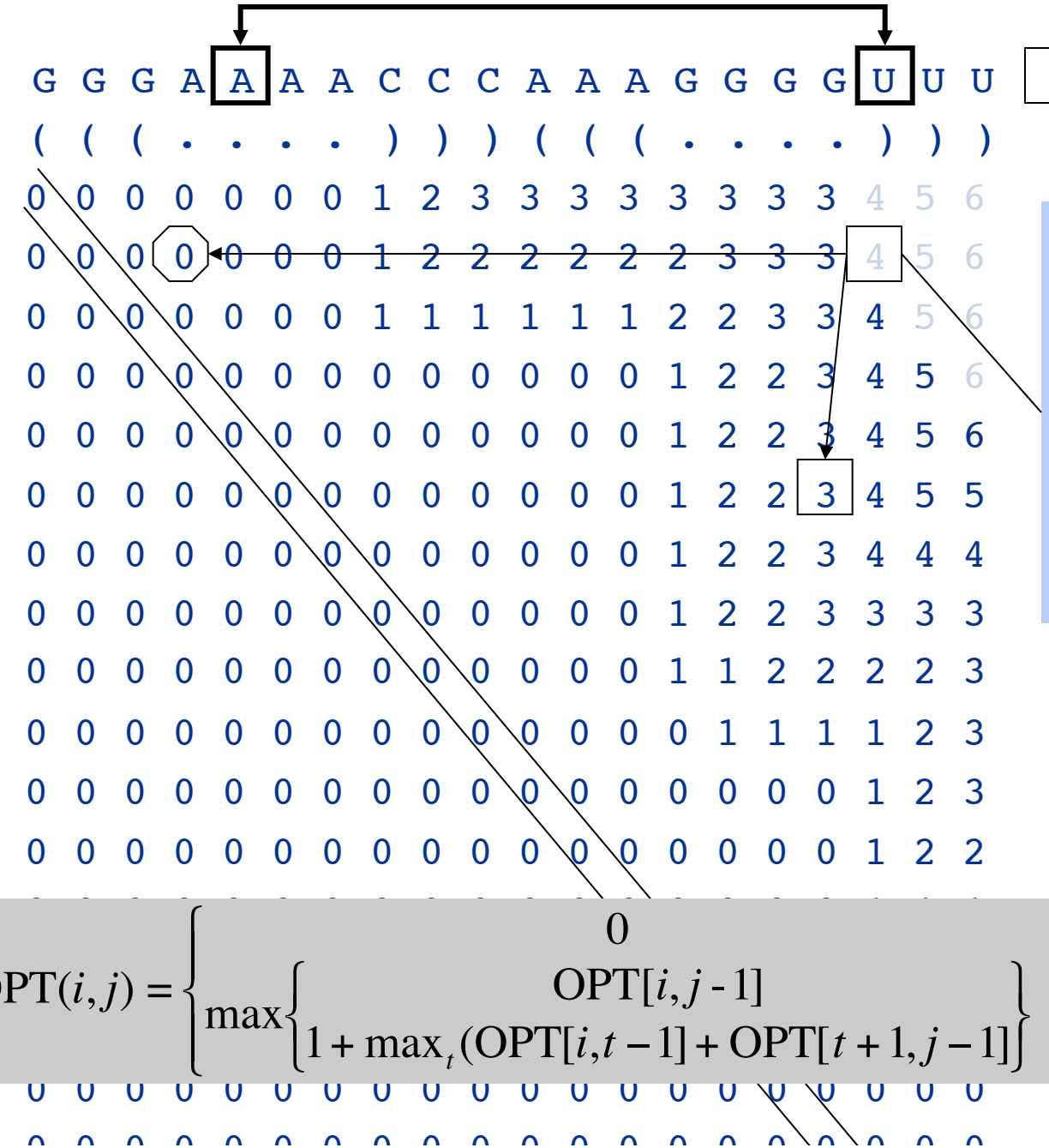


Case 3, 2 ≤ t <18-4:
t = 4: yes pair
OPT[2,18]≥1+0+3

GG**A**AAACCCAAAGGGG**U**
..(...(((....)))).

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

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

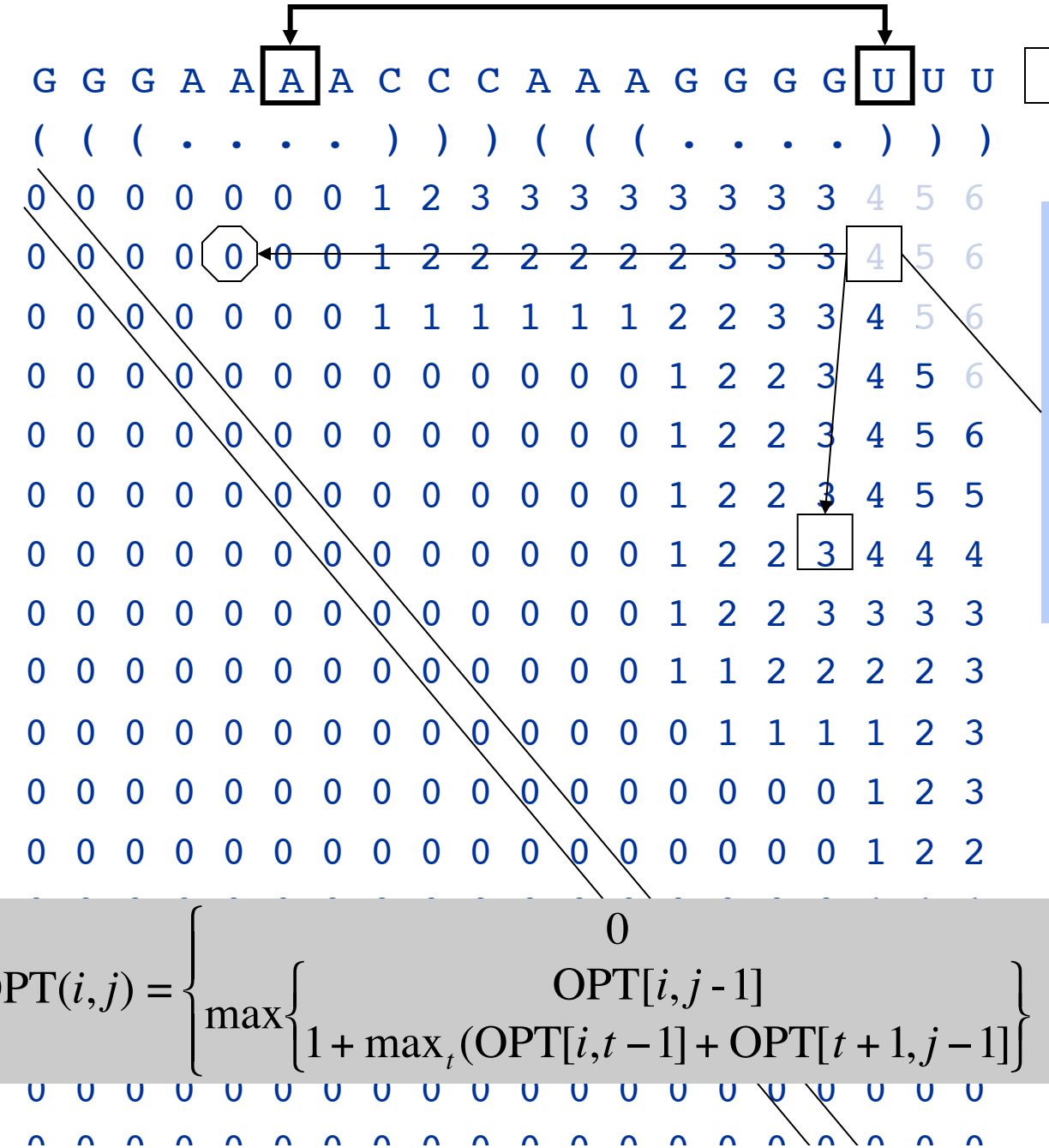


Case 3, 2 ≤ t <18-4:
t = 5: yes pair
OPT[2,18]≥1+0+3

GGA**A**AACCCAAAGGGG**U**
...(..(((....)))

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

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

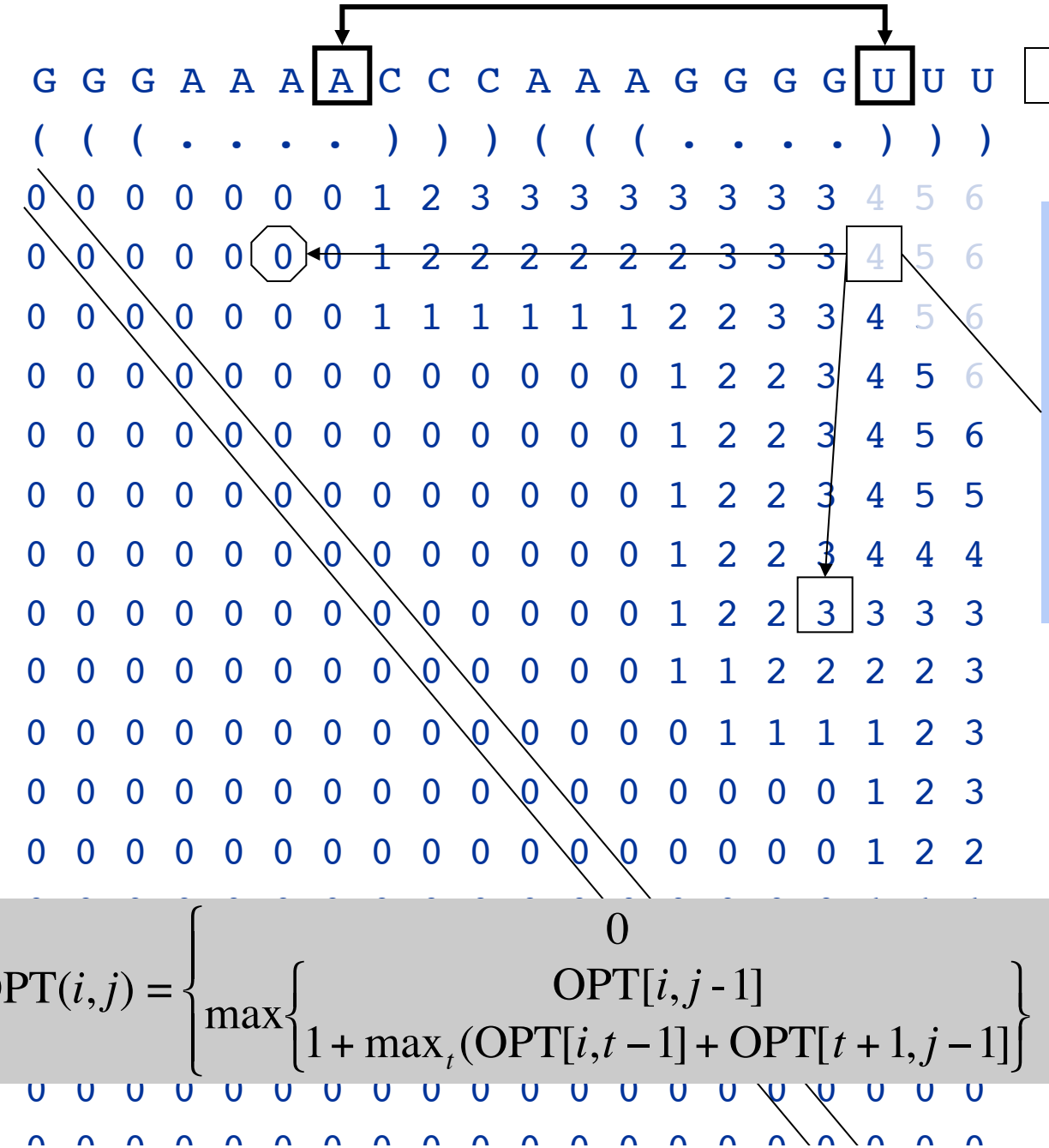


Case 3, $2 \le t < 18-4$:
t = 6: yes pair
OPT[2,18]≥1+0+3

GGAA**A**ACCCAAAGGGG**U**
....(.((((....))))

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

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

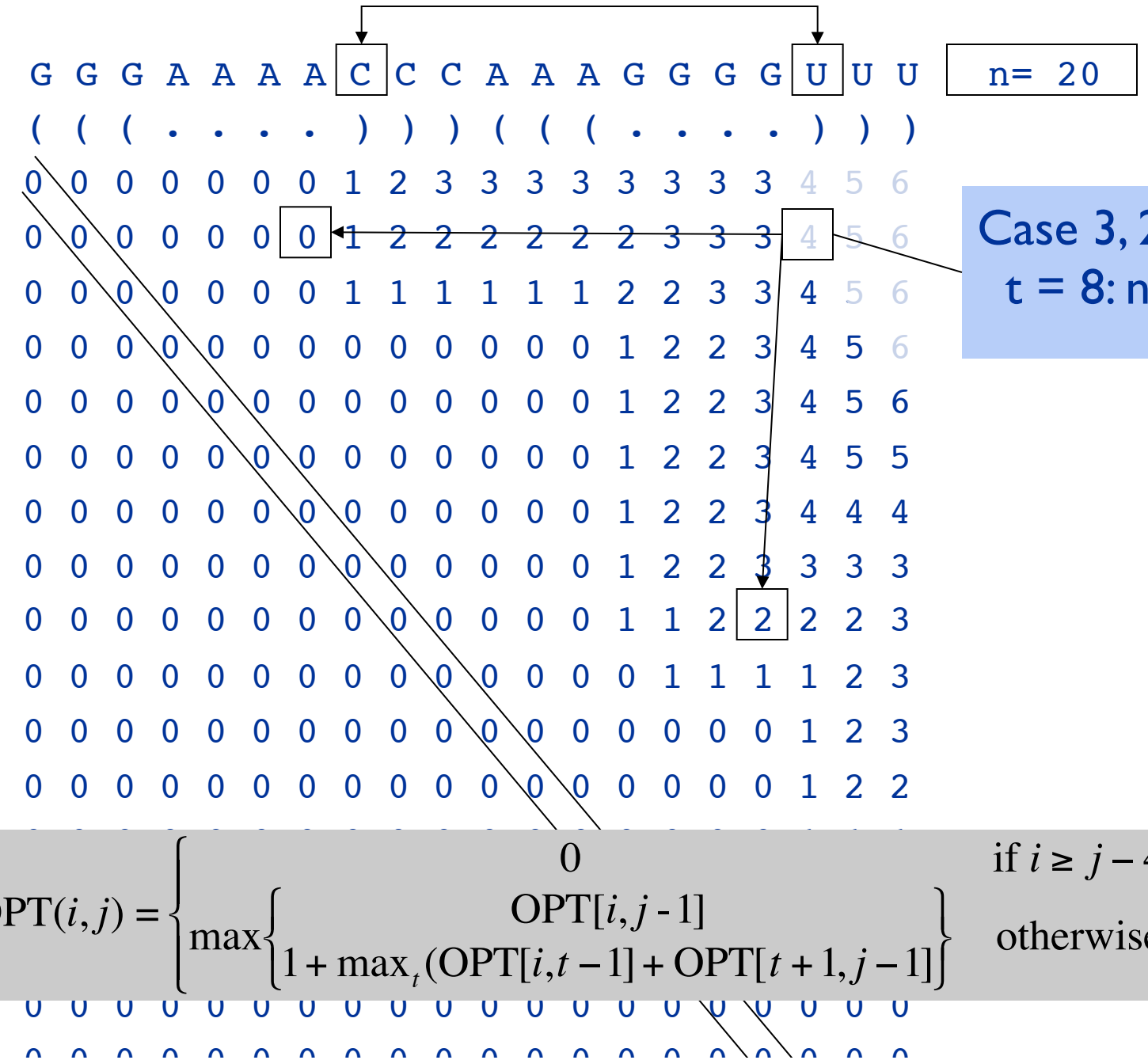


Case 3, $2 \le t < 18-4$:
t = 7: yes pair
OPT[2,18]≥1+0+3

GGAAA**A**CCCAAAGGGG**U**
.....((((....))))

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

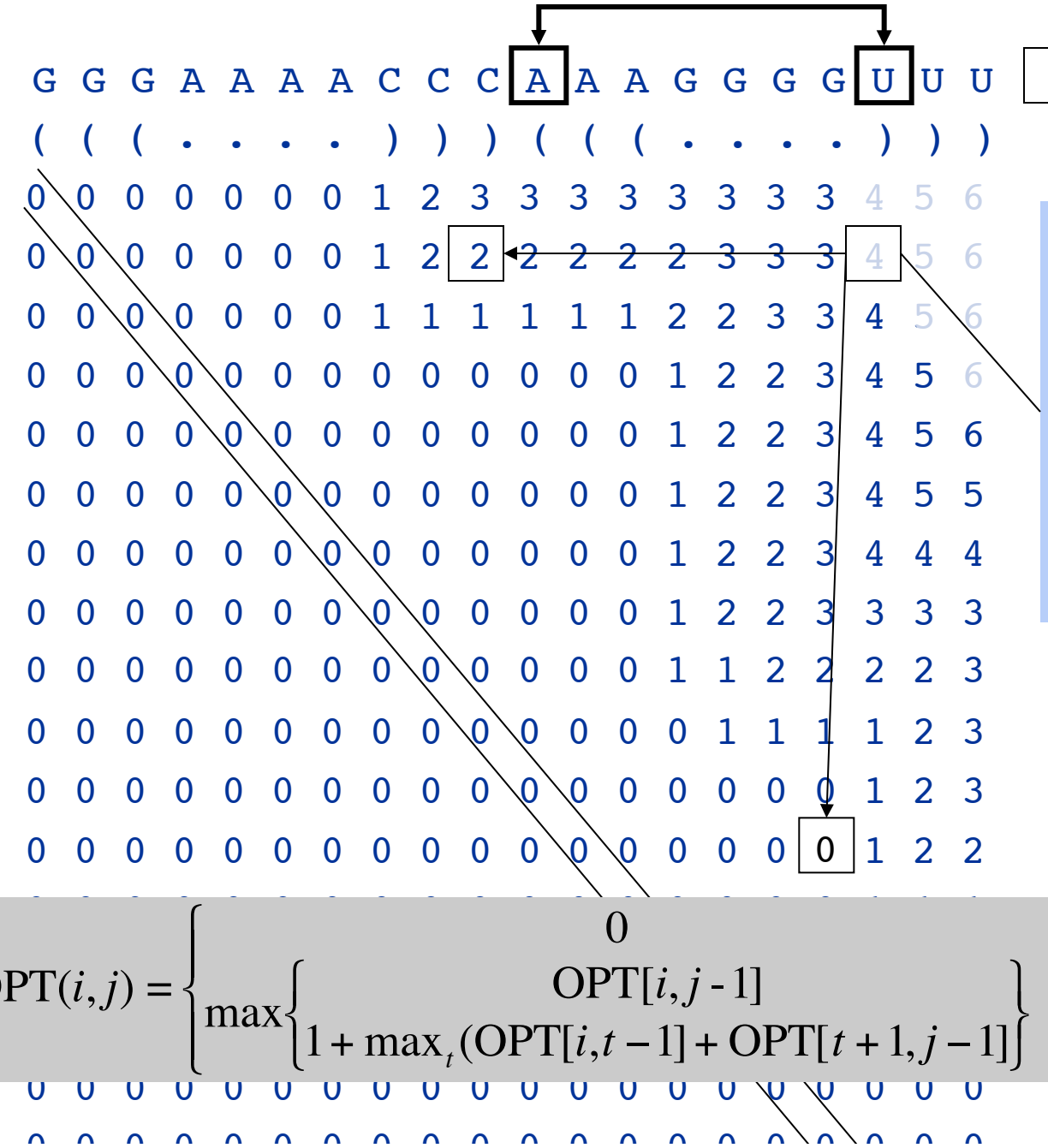
Computing one cell: OPT[2,18] = ?



Case 3, $2 \le t < 18-4$:
$t = 8$: no pair

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

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

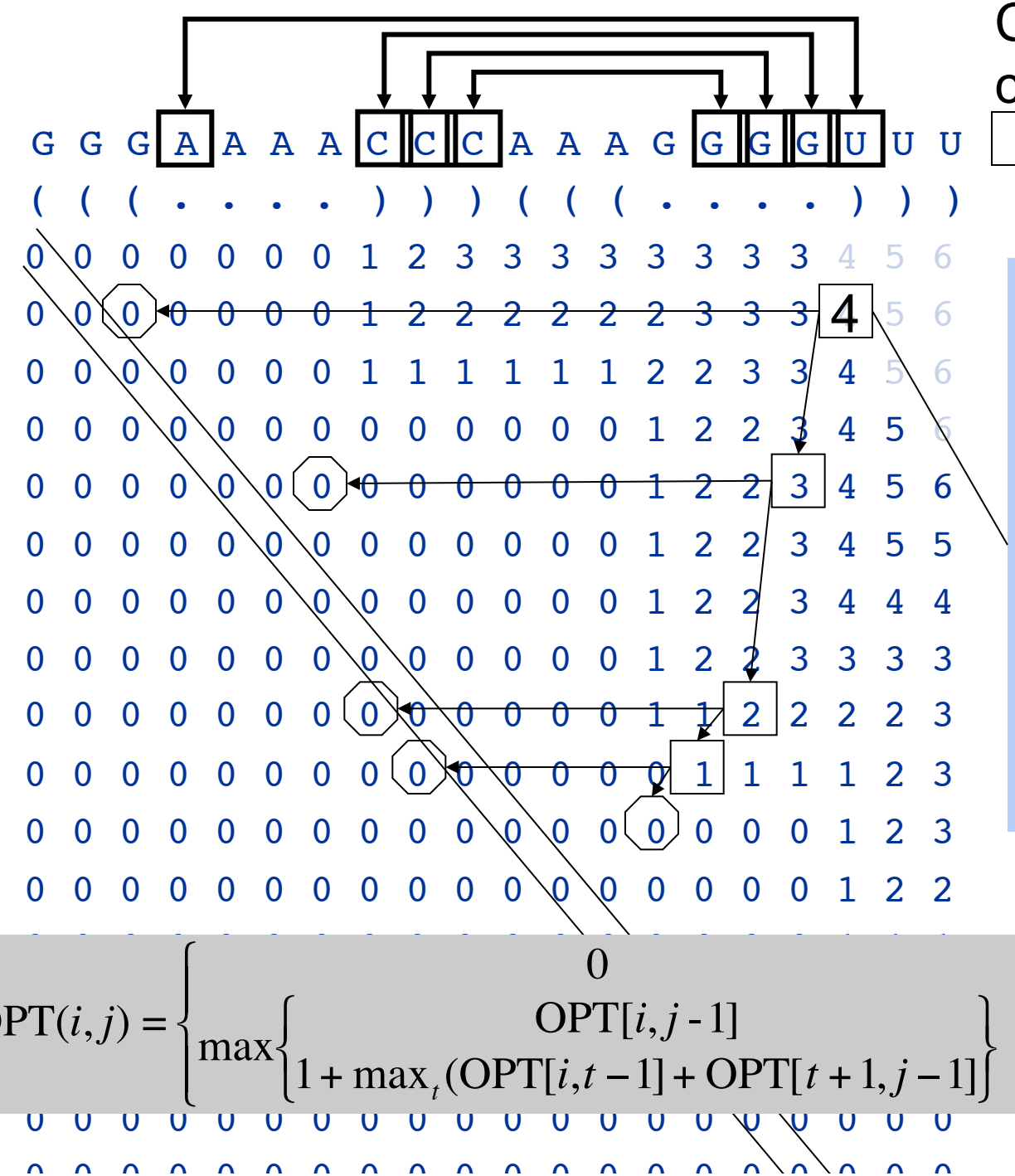


Case 3, $2 \le t < 18-4$:
t = 11: yes pair
OPT[2,18]≥1+2+0

GGAAAACCC**A**AAGGGG**U**
((.....))(.....)

(not shown:
t=9,10, 12,13)

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



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

n= 20

Overall, Max = 4 several ways, e.g.:

GGAAAACCCAAAGGGGU

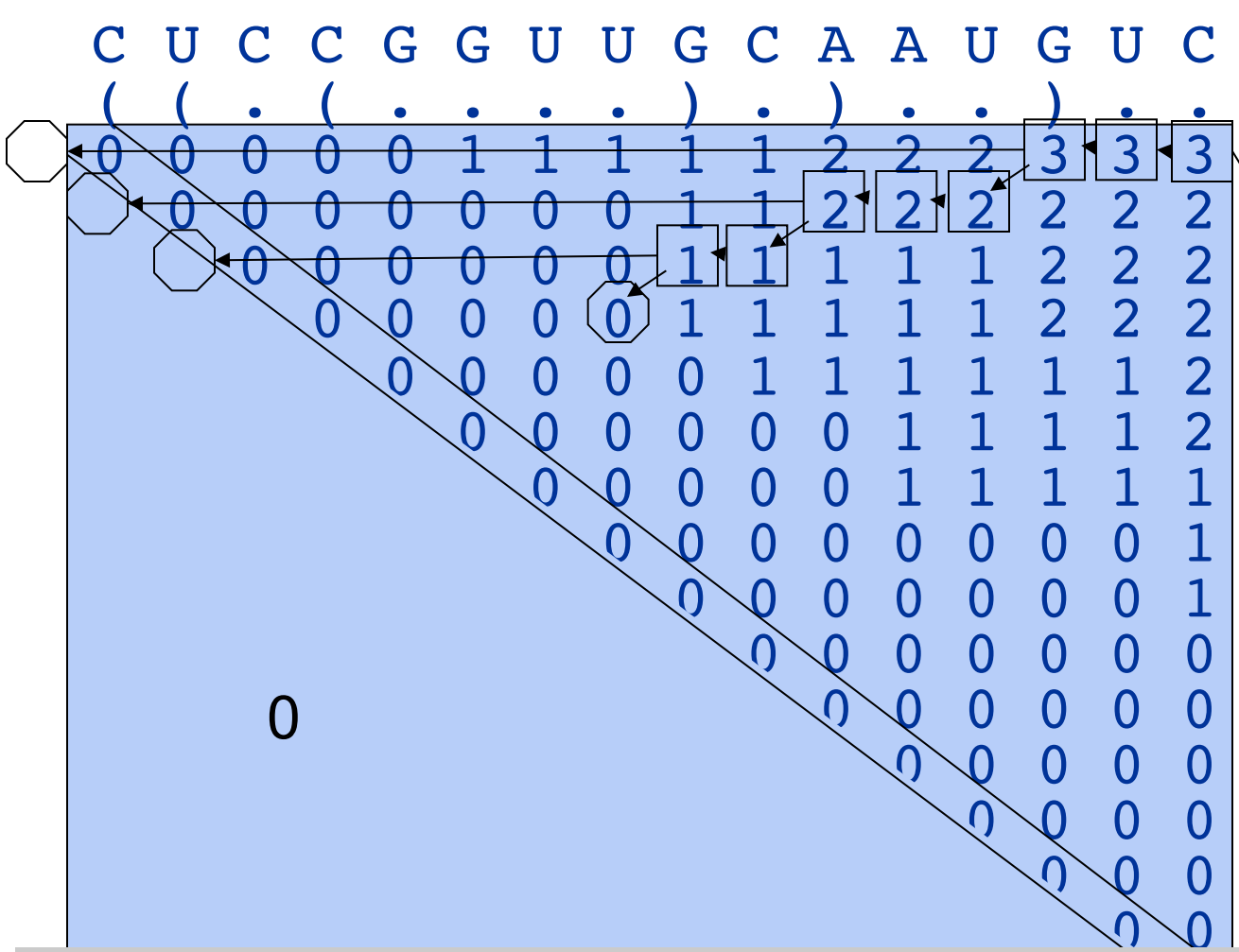
..(...(((....))))

tree shows trace back:

- square = case 3
- octagon = case 1

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

Another Trace Back Example



E.g.:
OPT[1,16] = 3:

CUCCGGUUGCAAUGUC

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

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

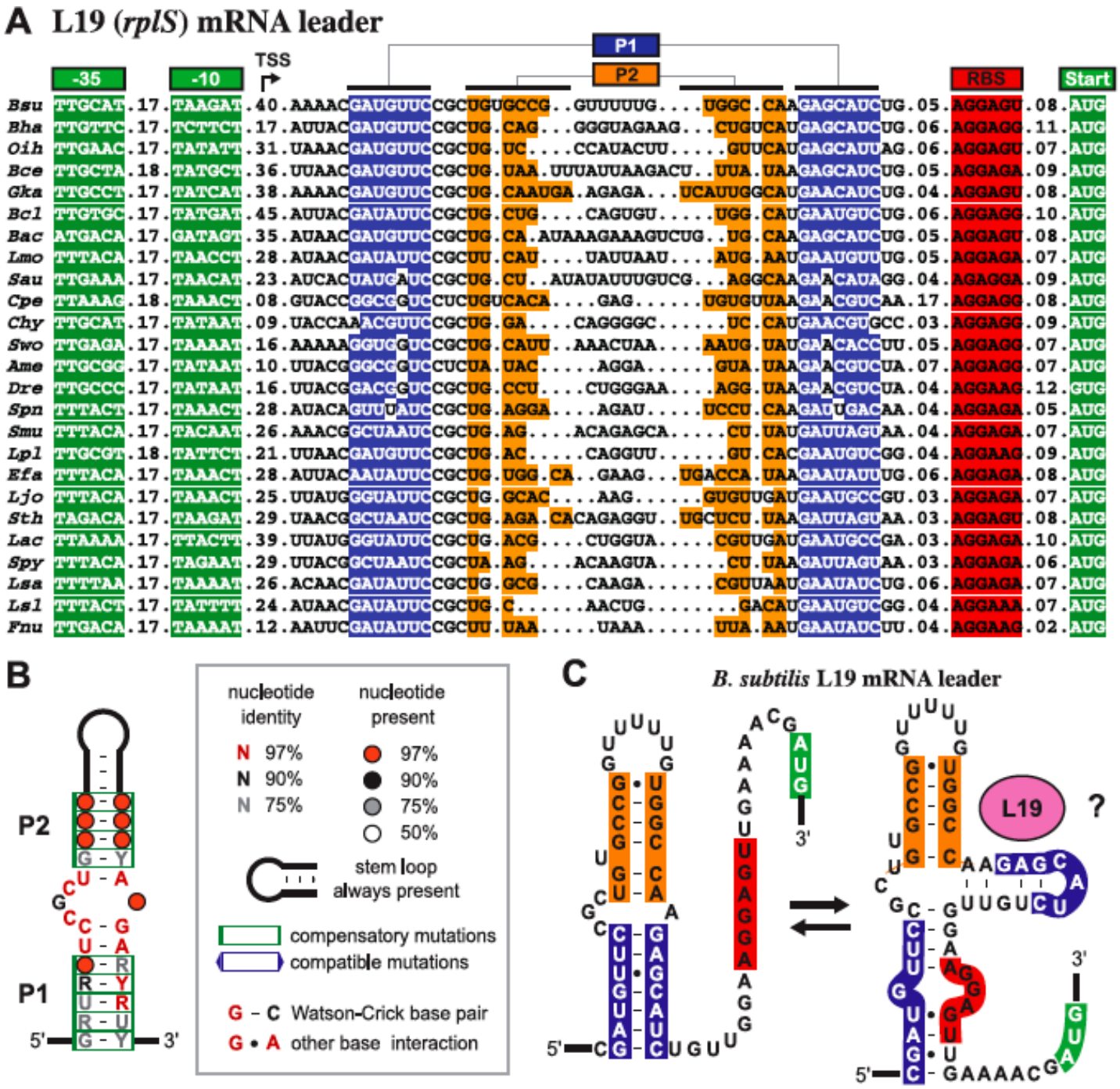


Figure 3. Putative Autoregulatory Structure in L19 mRNA Leaders