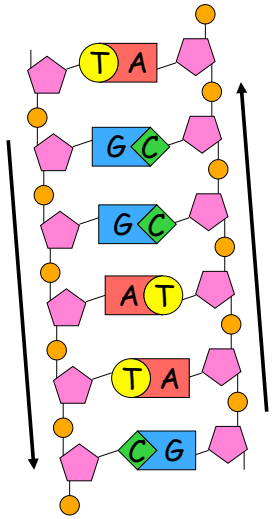
# RNA Secondary Structure



## DNA Structure

- Phosphodiester bonds (covalent) along backbone
- Hydrogen bonds between base pairs



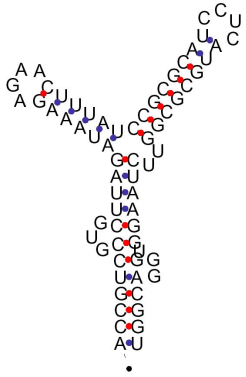



## RNA Secondary Structure

RNA primary sequence

ACCGUCCGUGCUUAGAUAAAGAGAACUUUAUCCGCGCAUCCUCAUGCGCGGUUCUAAGGUGGGACGGU

RNA secondary structure

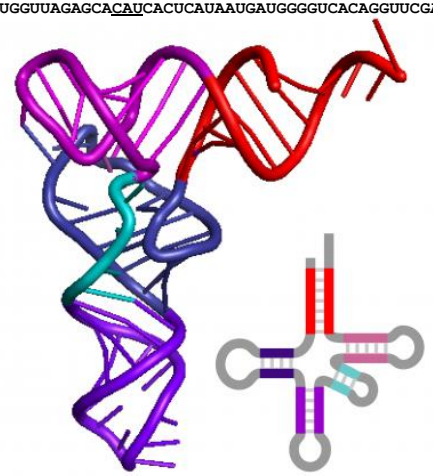




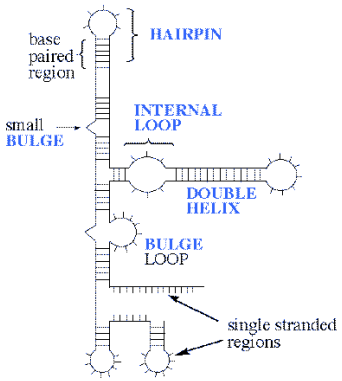
## tRNA Structure

tRNA primary sequence

GGCUACGUAGCUCAGUUGGUUAGAGCACAUCACUCAUAAUGAUGGGGUCACAGGUUCGAAUCCCGUCGUAGCCACC
A



## Components of RNA Secondary Structure



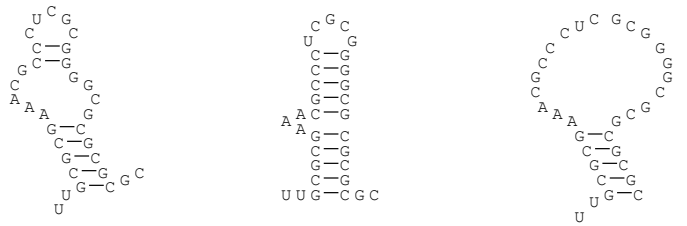



## Lots of Possible Structures for Each RNA Sequence

**RNA primary sequence**

UUGCGCGAAACGCCCUCGCGGGGCGCGCGCGC

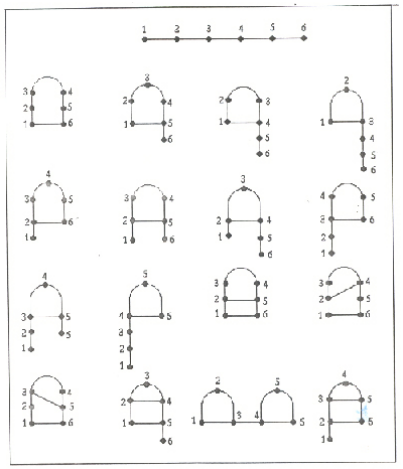
**Possible RNA secondary structures**





## Lots of Possible Structures for Each RNA Sequence

**Possible structures for a sequence of length $n$=6**

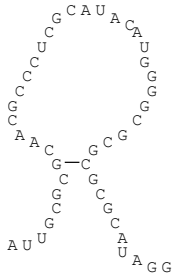


In general, the number of secondary structures $s(n)$ for a sequence of length $n$ is given by the following recurrence:

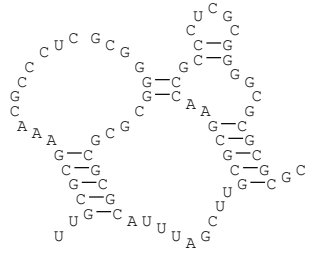
$$s(n) = \begin{cases} 1 & \text{if } n \le 2 \\ s(n-1) + \sum_{k=1}^{n-2} s(k-1) * s(n-k) & \text{if } n > 2 \end{cases}$$



## Disallow Pseudoknots



**Example pseudoknot**



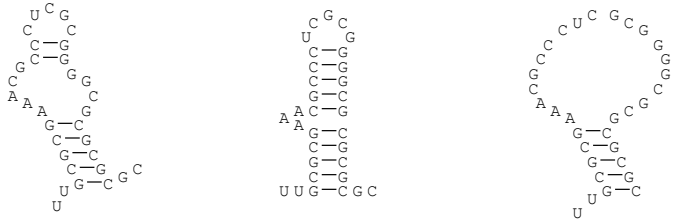
## Two Approaches

- Maximize the number of basepairs in the structure
- Minimize the energy of the structure

**RNA primary sequence**

UUGCGCGAAACGCCCUCGCGGGGCGCGCGCGC

**Possible RNA secondary structures**

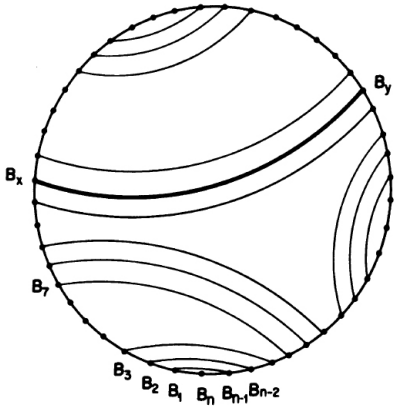


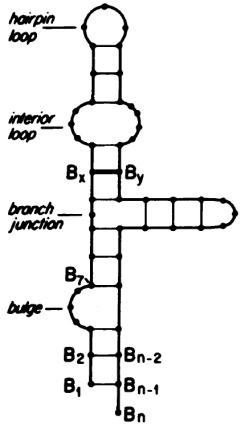


## Formulation as a Graph

**RNA primary sequence**

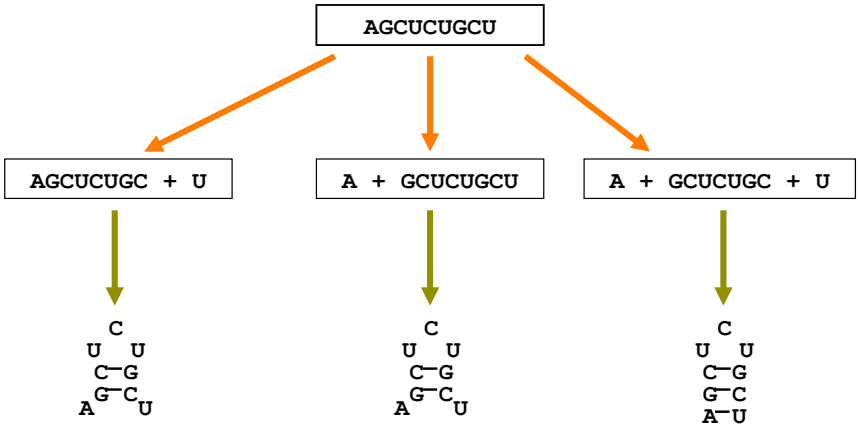
$B_1B_2B_3B_4B_5B_6B_7...B_{x-2}B_{x-1}B_xB_{x+1}B_{x+2}...B_{y-2}B_{y-1}B_yB_{y+1}B_{y+2}...B_{n-2}B_{n-1}B_n$





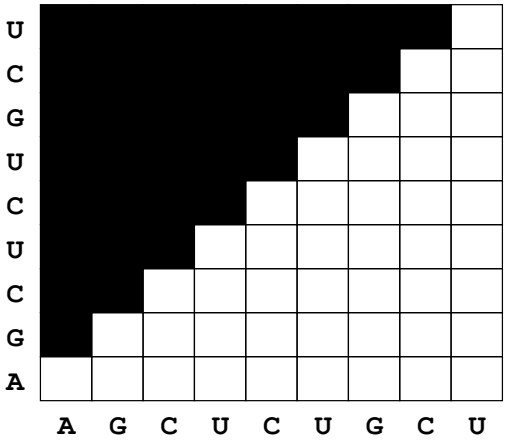

## Optimal Substructure






## Score Table

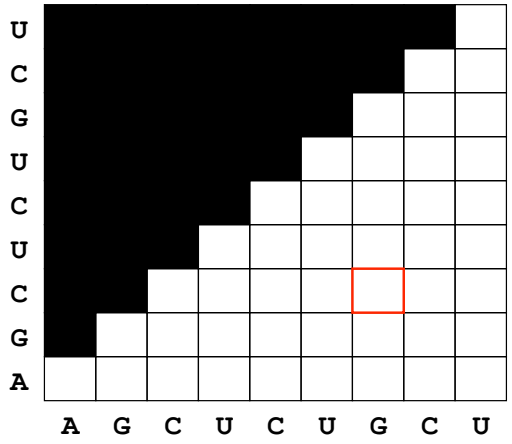
AGCUCUGCU



## Score Table

AGCUCUGCU

CUCUG

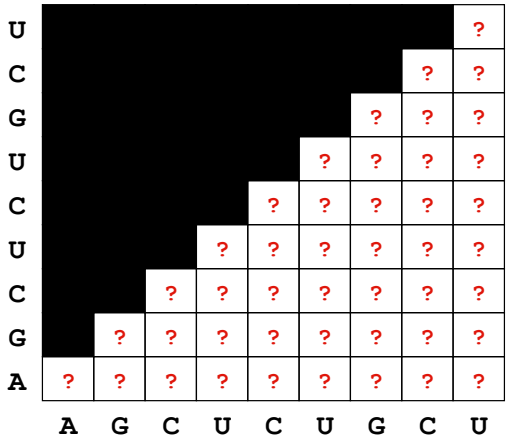





## Score Table

AGCUCUGCU

AGCUCUGCU






## Score Table

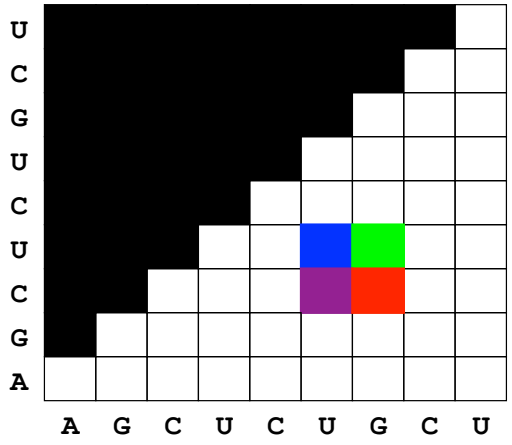
AGCUCUGCU

CUCU

CUCUG

UCU

UCUG

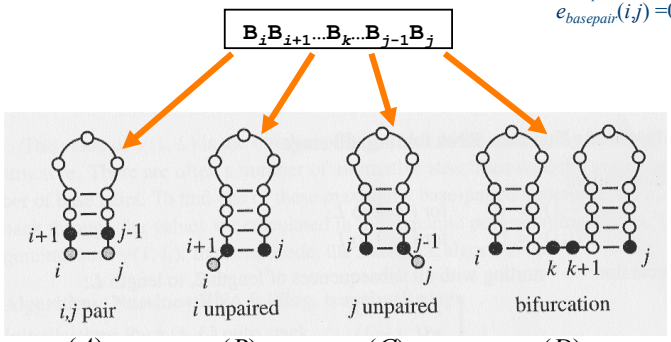





## Optimal Substructure

$$M(i,j) = \max \begin{cases} M(i+1,j-1) + e_{basepair}(i,j) & (A) \\ M(i+1,j) & (B) \\ M(i,j-1) & (C) \\ \max\limits_{i<k<j} \left( M(i,k) + M(k+1,j) \right) & (D) \end{cases}$$

$e_{basepair}(i,j)$ =1 if $i$ and $j$ pair
$e_{basepair}(i,j)$ =0 otherwise




(A) (B) (C) (D)

## Score Table

AGCUCUGCU

GCUCUGCU

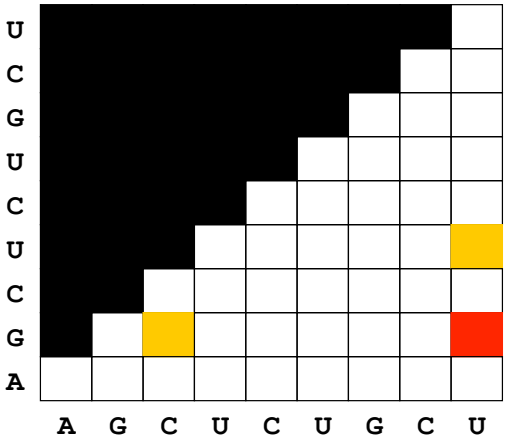
max

(A) CUCUGC + $e_{basepair}$(G,U)

(B) CUCUGCU

(C) GCUCUGC

(D) GC + UCUGCU






## Score Table

AGCUCUGCU

GCUCUGCU

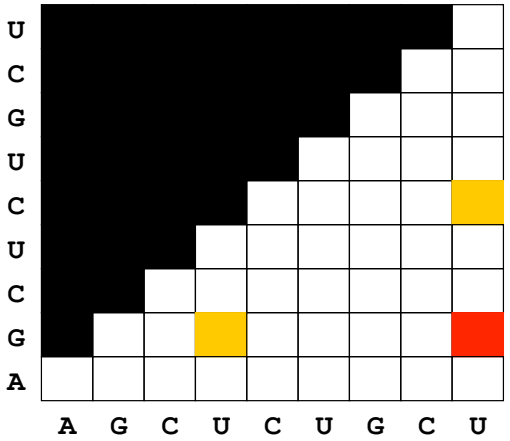
max

(A) CUCUGC + $e_{basepair}$(G,U)

(B) CUCUGCU

(C) GCUCUGC

(D) GCU + CUGCU






## Score Table

AGCUCUGCU

GCUCUGCU

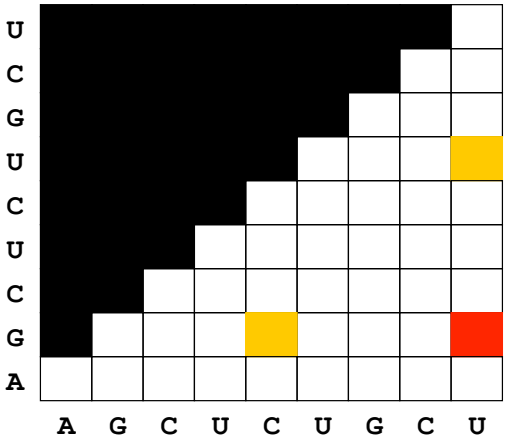
max

(A) CUCUGC + $e_{basepair}$(G,U)

(B) CUCUGCU

(C) GCUCUGC

(D) GCUC + UGCU






## Score Table

AGCUCUGCU

GCUCUGCU

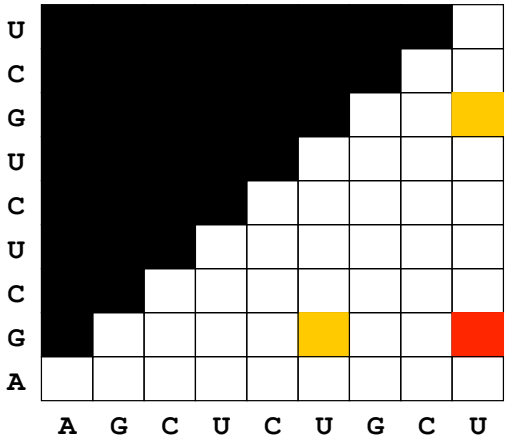
max

(A) CUCUGC + $e_{basepair}$(G,U)

(B) CUCUGCU

(C) GCUCUGC

(D) GCUCU + GCU



## Score Table

AGCUCUGCU

GCUCUGCU

max

(A) CUCUGC + $e_{basepair}$(G,U)

(B) CUCUGCU

(C) GCUCUGC

(D) GCUCUG + CU

| | A | G | C | U | C | U | G | C | U |
|---|---|---|---|---|---|---|---|---|---|
| U | | | | | | | | | |
| C | | | | | | | | | |
| G | | | | | | | | | |
| U | | | | | | | | | |
| C | | | | | | | | | |
| U | | | | | | | | | |
| C | | | | | | | | | |
| G | | | | | | | | | |
| A | | | | | | | | | |



## Score Table

AGCUCUGCU

| | A | G | C | U | C | U | G | C | U |
|---|---|---|---|---|---|---|---|---|---|
| U | | | | | | | | | 0 |
| C | | | | | | | | 0 | 0 |
| G | | | | | | | 0 | 0 | 0 |
| U | | | | | | 0 | 0 | 0 | 0 |
| C | | | | | 0 | 0 | 1 | 1 | 1 |
| U | | | | 0 | 0 | 0 | 1 | 1 | 1 |
| C | | | 0 | 0 | 0 | 0 | 1 | 1 | 1 |
| G | | 0 | 0 | 0 | 1 | 1 | 1 | 2 | 2 |
| A | 0 | 0 | 0 | 1 | 1 | 2 | 2 | 2 | 3 |



## Backtracking to Determine Structure

AGCUCUGCU

| | A | G | C | U | C | U | G | C | U |
|---|---|---|---|---|---|---|---|---|---|
| U | | | | | | | | | 0 |
| C | | | | | | | | 0 | 0 |
| G | | | | | | | 0 | 0 | 0 |
| U | | | | | | 0 | 0 | 0 | 0 |
| C | | | | | 0 | 0 | 1 | 1 | 1 |
| U | | | | 0 | 0 | 0 | 1 | 1 | 1 |
| C | | | 0 | 0 | 0 | 0 | 1 | 1 | 1 |
| G | | 0 | 0 | 0 | 1 | 1 | 1 | 2 | 2 |
| A | 0 | 0 | 0 | 1 | 1 | 2 | 2 | 2 | 3 |



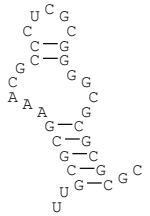
## Two Approaches

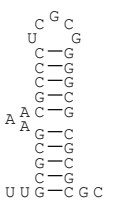
- Maximize the number of basepairs in the structure
- Minimize the energy of the structure

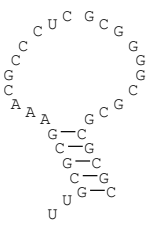
**RNA primary sequence**

UUGCGCGAAACGCCCUCGCGGGGCGCGCGCGC

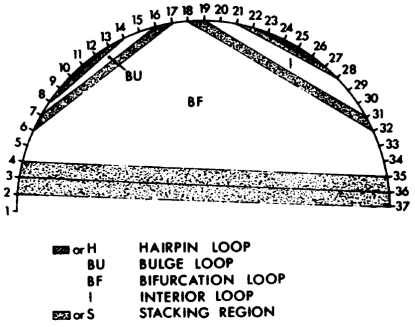
**Possible RNA secondary structures**

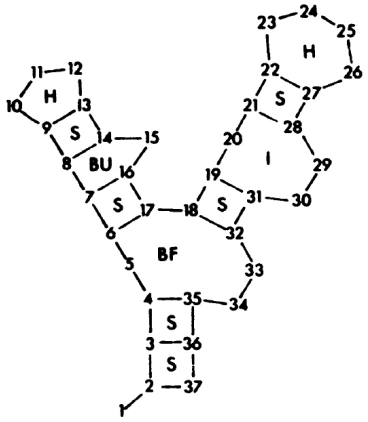






## Minimize Energy of Structure






## Energies of Components

| | Stacking Energies for base pairs | | | | | |
|---|---|---|---|---|---|---|
| | A/U | C/G | G/C | U/A | G/U | U/G |
| A/U | -0.9 | -1.8 | -2.3 | -1.1 | -1.1 | -0.8 |
| C/G | -1.7 | -2.9 | -3.4 | -2.3 | -2.1 | -1.4 |
| G/C | -2.1 | -2.0 | -2.9 | -1.8 | -1.9 | -1.2 |
| U/A | -0.9 | -1.7 | -2.1 | -0.9 | -1.0 | -0.5 |
| G/U | -0.5 | -1.2 | -1.4 | -0.8 | -0.4 | -0.2 |
| U/G | -1.0 | -1.9 | -2.1 | -1.1 | -1.5 | -0.4 |

| | Destabilizing Energies for Loops | | | | |
|---|---|---|---|---|---|
| Number of Bases | 1 | 5 | 10 | 20 | 30 |
| Internal | -- | 5.3 | 6.6 | 7.0 | 7.4 |
| Bulge | 3.9 | 4.8 | 5.5 | 6.3 | 6.7 |
| Hairpin | -- | 4.4 | 5.3 | 6.1 | 6.5 |



## Optimal Substructure

*V(i,j)* represents minimum energy structure of sequence between indices *i* and *j* such that the nucleotides at indices *i* and *j* basepair with each other

$$V(i,j) = \begin{cases} e_{hairpin}(i,j) & \text{// hairpin loop} \\ e_{stacking}(i,j,i+1,j-1) + V(i+1,j-1) & \text{// stacking region} \\ \min\limits_{i+1<i'<j'<j-1} \left(e_{interior}(i,j,i',j') + V(i',j')\right) & \text{// interior loop} \\ \min\limits_{i+1<j'<j-1} \left(e_{bulge}(i,j,i+1,j') + V(i+1,j')\right) & \text{// bulge loop (3')} \\ \min\limits_{i+1<i'<j-1} \left(e_{bulge}(i,j,i',j-1) + V(i',j-1)\right) & \text{// bulge loop (5')} \\ \min\limits_{i+1<i'<j-2} \left(W(i+1,i') + W(i'+1,j-1)\right) & \text{// bifurcation} \end{cases}$$

hairpin loop | stacking region | interior loop | bulge loop | bifurcation




## Optimal Substructure

*W(i,j)* represents minimum energy structure of sequence between indices *i* and *j* such that the nucleotides at indices *i* and *j* may or may not basepair with each other

$$W(i,j) = \begin{cases} W(i+1,j) & \text{// } i \text{ doesn't pair} \\ W(i,j-1) & \text{// } j \text{ doesn't pair} \\ V(i,j) & \text{// } i \text{ and } j \text{ pair with each other} \\ \min\limits_{i<i'<j-1} \left(W(i,i') + W(i'+1,j)\right) & \text{// } i \text{ and } j \text{ pair but not with each other} \end{cases}$$

*i* doesn't pair | *j* doesn't pair | *i* and *j* pair with each other | both *i* and *j* pair, but not with each other