



Bilkent University

IE400

Principles of Engineering Management

Project

Spring 2022-2023

Group 6

Team Members

Tuna Dalbeler - 21802539 - tuna.dalbeler@ug.bilkent.edu.tr

Efe Beydogan - 21901548 - efe.beydogan@ug.bilkent.edu.tr

Emir Melih Erdem - 21903100 - melih.erdem@ug.bilkent.edu.tr

Instructor: Professor Arnab Basu

May 5th, 2023

Table of Contents

Part A	3
Decision Variables:	3
Parameters:	3
Objective Function:	3
Results:	4
Part B	6
Decision Variables:	6
Parameters:	6
Objective Function:	6
Results:	7
Questions:	8
Part C	9
Decision Variables:	9
Parameters:	9
Objective Function:	9
Results:	10
Questions:	11
Part D	12
Decision Variables:	12
Parameters:	12
Objective Function:	12
Results:	13
Part E	15
Decision Variables:	15
Parameters:	15
Objective Function:	15
Results:	16
Part F	18
Parameters:	18
Dynamic Programming Model	18
Results:	19

***Note:** Throughout the report, the models are written using 1-based indexing (i.e. i and j start from 1 and go to n) whereas, in the Python code, 0-based indexing had to be used. Therefore, in the “Results” sections and the diagrams, the indices are 0-based.*

Part A

Decision Variables:

$$x_{ij} = \begin{cases} 1 & \text{if } S_i \text{ and } S_j \text{ is paired} \\ 0 & \text{otherwise} \end{cases}$$
$$1 \leq i, j \leq n$$

Parameters:

S_i : the i th element of nucleotide sequence S

n : length of the nucleotide sequence

Objective Function:

$$\text{maximize } \sum_{i=1}^n \sum_{j=1}^n x_{ij}$$

$$\text{subject to } \sum_{j=1}^n x_{ij} = 1, \forall i \in \{1, \dots, n\} \text{ (at most one pairing between elements)}$$

$$\sum_{i=1}^n x_{ij} = 1, \forall j \in \{1, \dots, n\} \text{ (at most one pairing between elements)}$$

$$x_{ij} + x_{jk} \leq 1, \forall i, j, k \in \{1, \dots, n\} \text{ (one nucleotide cannot be in two pairings)}$$

$$x_{ij} + x_{i'j'} \leq 1, i < i' < j < j', \forall i, j, i', j' \in \{1, \dots, n\} \text{ (pairings cannot cross each other)}$$

$$x_{ij} = 0, \text{ if } \begin{cases} S_i = A \text{ and } S_j \neq U \\ S_i = G \text{ and } S_j \neq C \\ S_i = U \text{ and } S_j \neq A \\ S_i = C \text{ and } S_j \neq G \end{cases} \quad \forall i, j \in \{1, \dots, n\}$$

(nucleotide A must be paired with only U and C must be paired with only G (or vice versa))

$$x_{ij} = 0, \text{ for all } j < i + 4, i = 1 \dots n, j = 1 \dots n$$

(Each nucleotide cannot pair with any nucleotide that is less than 4 positions away)

$$x_{ij} \in \{0, 1\}, \forall i, j \in \{1, \dots, n\}$$

Results:

Command line output:

Pairs: [(2, 62), (5, 61), (6, 60), (7, 59), (8, 58), (9, 56), (10, 55), (11, 54), (12, 53), (13, 42), (14, 22), (15, 21), (16, 20), (23, 41), (24, 40), (26, 37), (29, 36), (30, 34), (43, 51), (45, 50), (63, 69), (64, 68)]

Matches: [('A', 'U'), ('G', 'C'), ('G', 'C'), ('U', 'A'), ('A', 'U'), ('U', 'A'), ('A', 'U'), ('A', 'U'), ('G', 'C'), ('A', 'U'), ('C', 'G'), ('G', 'C'), ('A', 'U'), ('C', 'G'), ('C', 'G'), ('A', 'U'), ('U', 'A'), ('G', 'C'), ('U', 'A'), ('G', 'C'), ('G', 'C'), ('G', 'C')]

Number of pairs: 22

Optimal value: 22.0

Runtime: 13.23199987411499

Execution time: 813.5960848331451 seconds

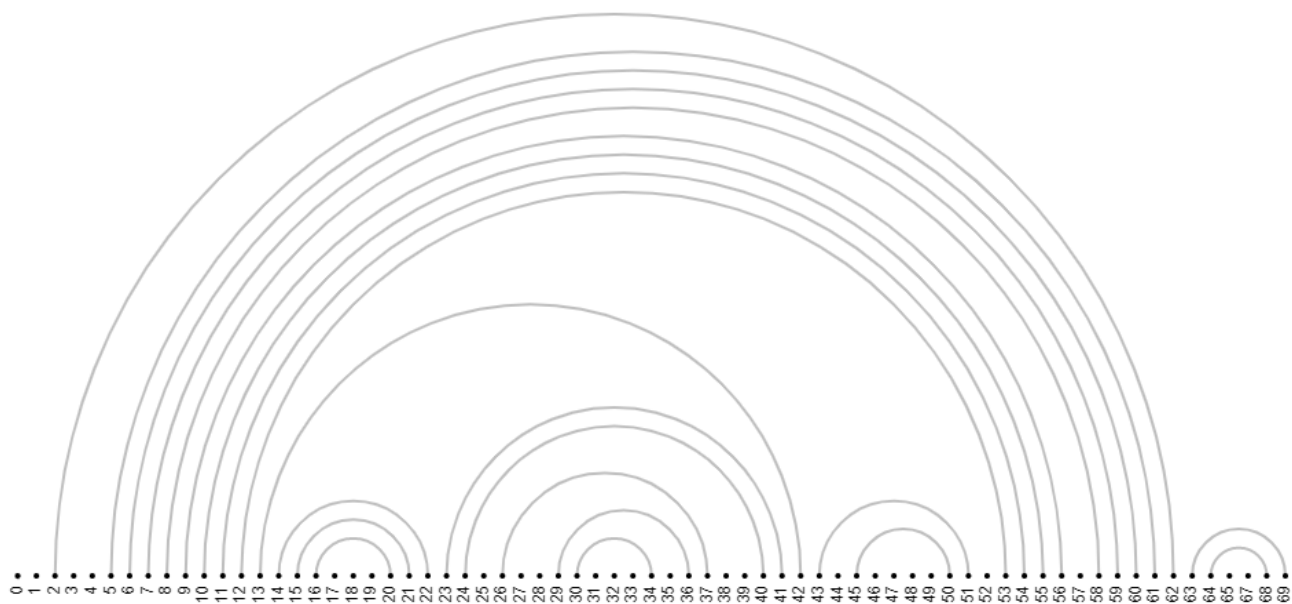


Figure 1: RNA matching diagram for Part A

Part B

Decision Variables:

$$x_{ij} = \begin{cases} 1 & \text{if } S_i \text{ and } S_j \text{ is paired} \\ 0 & \text{otherwise} \end{cases}$$
$$1 \leq i, j \leq n$$

Parameters:

S_i : the i th element of nucleotide sequence S

n : length of the nucleotide sequence

$E[f_1, f_2]$: Energy level of nucleotide pair (f_1, f_2)

Objective Function:

$$\text{Minimize } \sum_{i=1}^n \sum_{j=1}^n x_{ij} \cdot E[S_i, S_j]$$

$$\text{subject to } \sum_{j=1}^n x_{ij} = 1, \forall i \in \{1, \dots, n\} \text{ (at most one pairing between elements)}$$

$$\sum_{i=1}^n x_{ij} = 1, \forall j \in \{1, \dots, n\} \text{ (at most one pairing between elements)}$$

$$x_{ij} + x_{jk} \leq 1, \forall i, j, k \in \{1, \dots, n\} \text{ (one nucleotide cannot be in two pairings)}$$

$$x_{ij} + x_{i'j'} \leq 1, i < i' < j < j', \forall i, j, i', j' \in \{1, \dots, n\} \text{ (pairings cannot cross each other)}$$

$$x_{ij} = 0, \text{ if } \begin{cases} S_i = A \text{ and } S_j \neq U \\ S_i = G \text{ and } S_j \neq C \\ S_i = U \text{ and } S_j \neq A \\ S_i = C \text{ and } S_j \neq G \end{cases} \quad \forall i, j \in \{1, \dots, n\}$$

(nucleotide A must be paired with only U and C must be paired with only G (or vice versa))

$$x_{ij} = 0, \text{ for all } j < i + 4, i = 1 \dots n, j = 1 \dots n$$

(Each nucleotide cannot pair with any nucleotide that is less than 4 positions away)

$$x_{ij} \in \{0, 1\}, \forall i, j \in \{1, \dots, n\}$$

Results:

Command line output:

Pairs: [(1, 69), (2, 62), (4, 61), (5, 60), (7, 59), (8, 58), (9, 56), (10, 55), (11, 54), (12, 53), (14, 52), (15, 21), (16, 20), (22, 50), (23, 49), (24, 48), (25, 45), (26, 33), (27, 32), (34, 44), (36, 42), (64, 68)]

Matches: [('G', 'C'), ('A', 'U'), ('G', 'C'), ('G', 'C'), ('U', 'A'), ('A', 'U'), ('U', 'A'), ('A', 'U'), ('A', 'U'), ('G', 'C'), ('C', 'G'), ('G', 'C'), ('A', 'U'), ('G', 'C'), ('C', 'G'), ('C', 'G'), ('C', 'G'), ('A', 'U'), ('G', 'C'), ('C', 'G'), ('A', 'U'), ('G', 'C')]

Number of pairs: 22

Optimal value: -30.82

Runtime: 27.391000032424927

Execution time: 830.2231049537659 seconds

Questions:

What are the differences in terms of solution time and the number of pairs?

	Part A	Part B
solution time	13.2 Seconds	27.39 Seconds
number of pairs	22	22

Although the number of pairs are the same, in Part B, the solution time is significantly longer and the matched pairs are different. For Part B, the optimizer tried to form pairs between C-G more than A-U, as the energy level of C-G is lower than A-U, so C-G pairs contribute more to minimizing the total free energy. The optimizer takes longer to find a solution in Part B as now the optimizer needs to explore more options, even when the maximum number of pairs has been reached, in order to find the pairing with the minimum total energy.

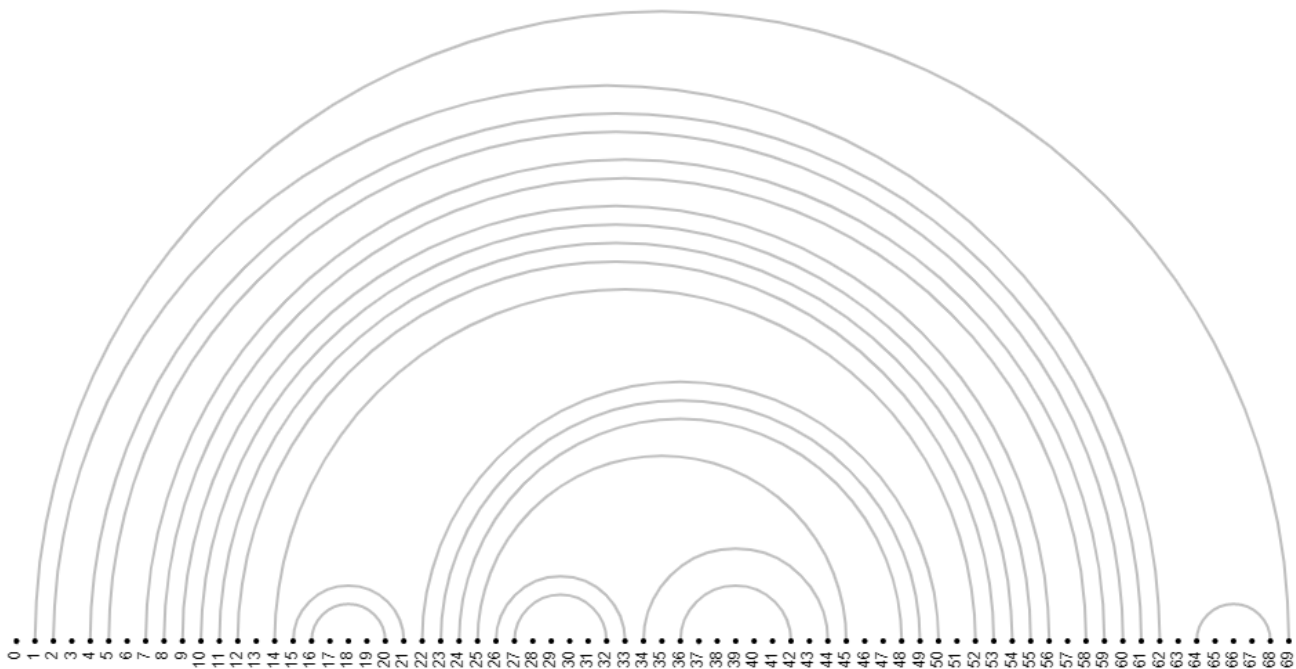


Figure 2: RNA matching diagram for Part B

Part C

Decision Variables:

$$x_{ij} = \begin{cases} 1 & \text{if } S_i \text{ and } S_j \text{ is paired} \\ 0 & \text{otherwise} \end{cases}$$
$$1 \leq i, j \leq n$$

Parameters:

S_i : the i th element of nucleotide sequence S

n : length of the nucleotide sequence

$E[f_1, f_2]$: Energy level of nucleotide pair (f_1, f_2)

Objective Function:

$$\text{Minimize } \sum_{i=1}^n \sum_{j=1}^n x_{ij} \cdot E[S_i, S_j]$$

$$\text{subject to } \sum_{j=1}^n x_{ij} = 1, \forall i \in \{1, \dots, n\} \text{ (at most one pairing between elements)}$$

$$\sum_{i=1}^n x_{ij} = 1, \forall j \in \{1, \dots, n\} \text{ (at most one pairing between elements)}$$

$$x_{ij} + x_{jk} \leq 1, \forall i, j, k \in \{1, \dots, n\} \text{ (one nucleotide cannot be in two pairings)}$$

$$x_{ij} + x_{i'j'} \leq 1, i < i' < j < j', \forall i, j, i', j' \in \{1, \dots, n\} \text{ (pairings cannot cross each other)}$$

$$x_{ij} = 0, \text{ if } \begin{cases} S_i = A \text{ and } S_j \neq U \\ S_i = G \text{ and } S_j \neq C \\ S_i = U \text{ and } S_j \neq A \\ S_i = C \text{ and } S_j \neq G \end{cases} \quad \forall i, j \in \{1, \dots, n\}$$

(nucleotide A must be paired with only U and C must be paired with only G (or vice versa))

Distance constraint: (Updated from the constraint from part a and b)

$$x_{ij} = 0, \text{ for all } j < i + 7, i = 1 \dots n, j = 1 \dots n$$

(Each nucleotide cannot pair with any nucleotide that is less than 7 positions away)

$$x_{ij} \in \{0, 1\}, \forall i, j \in \{1, \dots, n\}$$

Results:

Command line output:

Pairs: [(1, 69), (4, 68), (5, 61), (6, 60), (7, 59), (8, 58), (9, 56), (10, 55), (11, 54), (12, 53), (14, 52), (15, 50), (16, 47), (17, 46), (21, 45), (23, 44), (24, 41), (25, 40), (26, 39), (27, 34)]

Matches: [('G', 'C'), ('G', 'C'), ('G', 'C'), ('G', 'C'), ('U', 'A'), ('A', 'U'), ('U', 'A'), ('A', 'U'), ('A', 'U'), ('G', 'C'), ('C', 'G'), ('G', 'C'), ('A', 'U'), ('A', 'U'), ('C', 'G'), ('C', 'G'), ('C', 'G'), ('C', 'G'), ('A', 'U'), ('G', 'C')]

Number of pairs: 20

Optimal value: -28.040000000000006

Runtime: 12.097000122070312

Execution time: 805.5604355335236 seconds

Questions:

What are the differences in terms of solution time and the number of pairs?

	Part A	Part B	Part C
solution time	13.2 Seconds	18.6 Seconds	12.09 Seconds
number of pairs	22	22	20

Since the updated constraint is more restrictive than the former, there are less possible pairings and, therefore, the feasible region is narrowed down. With fewer solutions to explore, Part C takes shorter than both A and B. There are less pairs in the solution because there are less possible pairs to match due to the increased distance constraint between pairs.

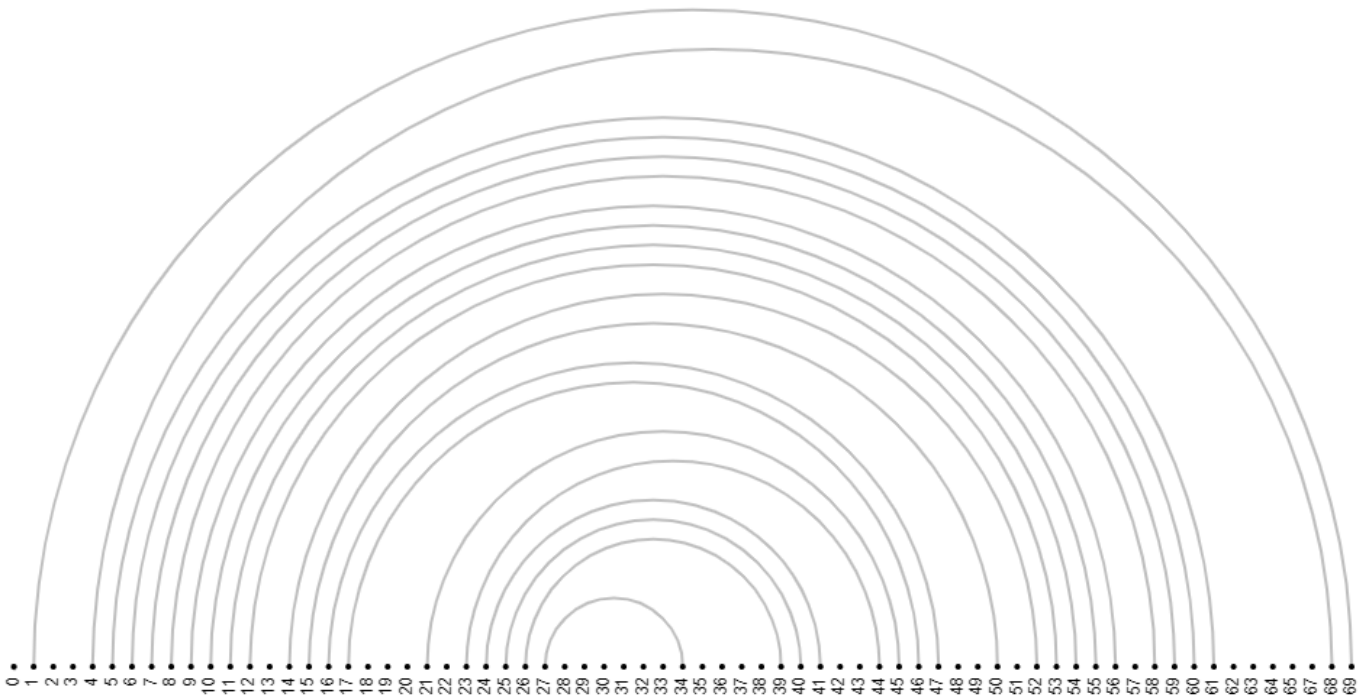


Figure 3: RNA matching diagram for Part C

Part D

Decision Variables:

$$x_{ij} = \begin{cases} 1 & \text{if } S_i \text{ and } S_j \text{ is paired} \\ 0 & \text{otherwise} \end{cases}$$

$$1 \leq i, j \leq N$$

$$y_{ij} = \begin{cases} 1 & \text{if } x_{ij} = 1 \text{ and } x_{i+1, j-1} = 1 \\ 0 & \text{otherwise} \end{cases}$$

$$1 \leq i, j \leq N$$

Parameters:

S_i : the i th element of nucleotide sequence S

n : length of the nucleotide sequence

$E[(f_{11}, f_{12}), (f_{21}, f_{22})]$: Energy level of stack pair $((f_{11}, f_{12}), (f_{21}, f_{22}))$

Objective Function:

$$\text{Minimize } \sum_{i=1}^n \sum_{j=1}^n y_{ij} \cdot E[(S_i, S_j), (S_{i+1}, S_{j-1})]$$

$$\text{subject to } \sum_{j=1}^n x_{ij} = 1, \forall i \in \{1, \dots, n\} \text{ (at most one pairing between elements)}$$

$$\sum_{i=1}^n x_{ij} = 1, \forall j \in \{1, \dots, n\} \text{ (at most one pairing between elements)}$$

$$x_{ij} + x_{jk} \leq 1, \forall i, j, k \in \{1, \dots, n\} \text{ (one nucleotide cannot be in two pairings)}$$

$$x_{ij} + x_{i'j'} \leq 1, i < i' < j < j', \forall i, j, i', j' \in \{1, \dots, n\} \text{ (pairings cannot cross each other)}$$

$$x_{ij} = 0, \text{ if } \begin{cases} S_i = A \text{ and } S_j \neq U \\ S_i = G \text{ and } S_j \neq C \\ S_i = U \text{ and } S_j \neq A \\ S_i = C \text{ and } S_j \neq G \end{cases} \quad \forall i, j \in \{1, \dots, n\}$$

(nucleotide A must be paired with only U and C must be paired with only G (or vice versa))

$$x_{ij} = 0, \text{ for all } j < i + 4, i = 1 \dots n, j = 1 \dots n$$

(Each nucleotide cannot pair with any nucleotide that is less than 4 positions away)

New constraint: to ensure there is a stack only when necessary pairs are present

$$2 \cdot y_{ij} \leq x_{ij} + x_{i+1, j-1} \text{ for } \forall i, j = 1 \dots n \text{ (nested pairing constraint)}$$

$$x_{ij} \in \{0, 1\}, \forall i, j \in \{1, \dots, n\}$$

$$y_{ij} \in \{0, 1\}, \forall i, j \in \{1, \dots, n\}$$

Results:

Command line output:

Pairs: [(1, 34), (2, 33), (3, 26), (4, 25), (5, 24), (6, 23), (14, 22), (15, 21), (16, 20), (40, 69), (41, 68), (44, 61), (45, 60), (46, 59), (51, 58), (52, 57)]

Matches: [('G', 'C'), ('A', 'U'), ('U', 'A'), ('G', 'C'), ('G', 'C'), ('G', 'C'), ('C', 'G'), ('G', 'C'), ('A', 'U'), ('G', 'C'), ('G', 'C'), ('G', 'C'), ('G', 'C'), ('U', 'A'), ('A', 'U'), ('G', 'C')]

Number of pairs: 16

Optimal value: -26.5

Runtime: 1.6010000705718994

Execution time: 799.8012628555298 seconds

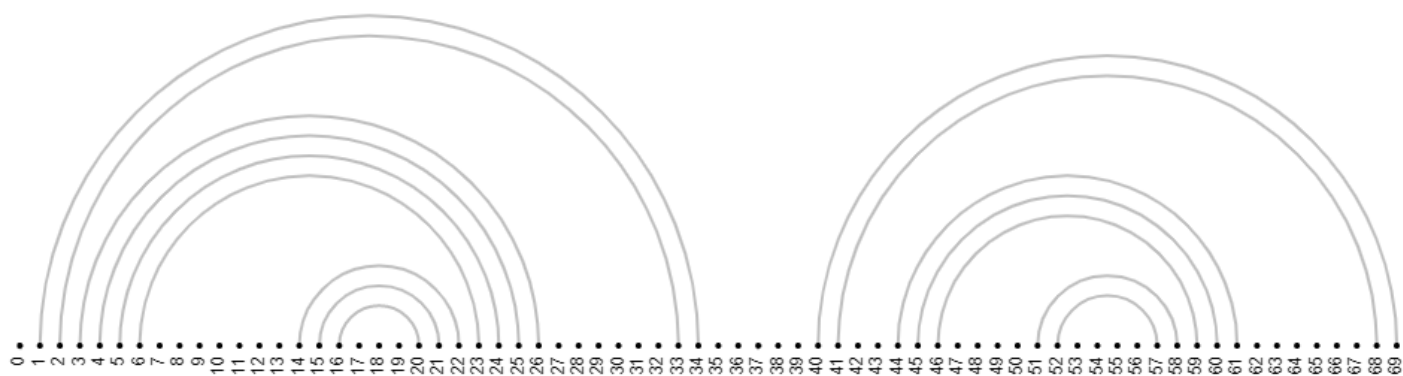


Figure 4: RNA matching diagram for Part D

Part E

Decision Variables:

$$x_{ijk} = \begin{cases} 1 & \text{if } S_i \text{ and } S_j \text{ is paired on side } k \\ 0 & \text{otherwise} \end{cases}$$

$$1 \leq i, j \leq N, \quad k \in \{0, 1\} \quad (k \text{ indicates pairing above or below})$$

$$y_{ij} = \begin{cases} 1 & \text{if } x_{ij} = 1 \text{ and } x_{i+1, j-1} = 1 \\ 0 & \text{otherwise} \end{cases}$$

$$1 \leq i, j \leq N$$

Parameters:

S_i : the i th element of nucleotide sequence S

n : length of the nucleotide sequence

$E[(f_{11}, f_{12}), (f_{21}, f_{22})]$: Energy level of stack pair $((f_{11}, f_{12}), (f_{21}, f_{22}))$

Objective Function:

$$\text{Minimize } \sum_{i=1}^n \sum_{j=1}^n y_{ij} \cdot E[(S_i, S_j), (S_{i+1}, S_{j-1})]$$

subject to $x_{ij,0} + x_{j,i',0} + x_{ij,1} + x_{j,i',1} \leq 1, \quad \forall i, j, i' \in \{1 \dots n\}$ (one nucleotide cannot be in two pairings)

$$\sum_{i=1}^n \sum_{k \in \{0,1\}} x_{ijk} = 1, \quad \forall j \in \{1, \dots, n\} \quad (\text{at most one pairing})$$

$$\sum_{j=1}^n \sum_{k \in \{0,1\}} x_{ijk} = 1, \quad \forall i \in \{1, \dots, n\} \quad (\text{at most one pairing})$$

$$x_{ijk} = \begin{cases} S_i = A \text{ and } S_j \neq U \\ S_i = G \text{ and } S_j \neq C \\ S_i = U \text{ and } S_j \neq A \\ S_i = C \text{ and } S_j \neq G \end{cases} \quad \forall i, j \in \{1, \dots, n\}, \forall k \in \{0, 1\}$$

(nucleotide A must be paired with only U and C must be paired with only G (or vice versa))

$$x_{ijk} + x_{i'j'k} \leq 1, \forall i, j, i', j' \in \{1 \dots n\} \text{ and } \forall k \in \{0, 1\}$$

(Overlapping Constraint)

$$2 \cdot y_{ij} \leq x_{ij0} + x_{i+1j-10} + x_{ij1} + x_{i+1j-11}, \forall i, j \in \{1 \dots n\} \text{ (nested pairing constraint)}$$

$$x_{ijk} = 0, j < i + 4, \forall i, j \in \{1, \dots, n\}, \forall k \in \{0, 1\}$$

(each nucleotide cannot pair with any nucleotide that is less than 4 positions away)

$$x_{ijk} \in \{0, 1\}, \forall i, j \in \{1, \dots, n\}, \forall k \in \{0, 1\}$$

$$y_{ij} \in \{0, 1\}, \forall i, j \in \{1, \dots, n\}$$

Results:

Command line output:

Pairs: [(1, 34), (2, 33), (3, 26), (4, 25), (5, 24), (6, 23), (10, 66), (11, 65), (14, 22), (15, 21), (16, 20), (17, 58), (18, 57), (19, 56), (31, 53), (32, 52), (40, 61), (41, 60), (42, 59), (43, 51), (44, 50), (63, 69), (64, 68)]

Matches: [('G', 'C'), ('A', 'U'), ('U', 'A'), ('G', 'C'), ('G', 'C'), ('G', 'C'), ('A', 'U'), ('A', 'U'), ('C', 'G'), ('G', 'C'), ('A', 'U'), ('A', 'U'), ('G', 'C'), ('U', 'A'), ('G', 'C'), ('C', 'G'), ('G', 'C'), ('G', 'C'), ('U', 'A'), ('U', 'A'), ('G', 'C'), ('G', 'C'), ('G', 'C')]

Above/below: ['below', 'below', 'below', 'below', 'below', 'below', 'below', 'above', 'above', 'below', 'below', 'below', 'above', 'above', 'above', 'above', 'above', 'below', 'below', 'below', 'below', 'below', 'below', 'below']

Number of pairs: 23

Optimal value: -33.9

Runtime: 7.006999969482422

Execution time: 1738.668952703476 seconds

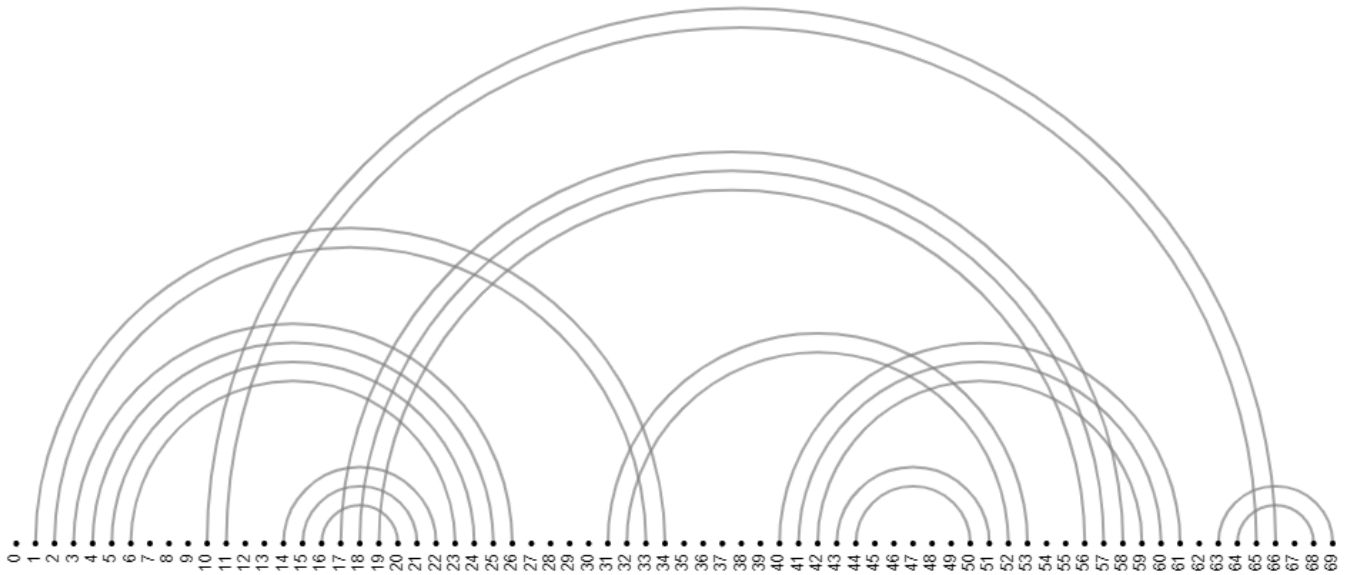


Figure 5: RNA matching diagram for Part E

Part F

Parameters:

S_i : the i th element of nucleotide sequence S

n : length of the nucleotide sequence

$SPE[(f_{11}, f_{12}), (f_{21}, f_{22})]$: Energy level of stack pair $((f_{11}, f_{12}), (f_{21}, f_{22}))$

Dynamic Programming Model

$e[i, j]$: min total energy constructed from the substring from i to j

$$e[i, j] = \begin{cases} 0 & \text{if } i > j - 4 \\ \min \left(\min_{i \leq k < j} (e[i, k] + e[k + 1, j]), b[i, j] \right) & \text{if } (S_i, S_j) \text{ is matching} \\ \min_{i \leq k < j} (e[i, k] + e[k + 1, j]) & \text{otherwise} \end{cases}$$

$b[i, j]$: min total energy constructed from the substring from i to j when (i, j) is paired

$$b[i, j] = \begin{cases} 0 & \text{if } i > j - 4 \text{ or } (S_i, S_j) \text{ is not matching} \\ SPE[(i, j), (i + 1, j - 1)] + b[i + 1, j - 1] & \text{if } i \leq j - 6 \text{ and } (S_{i+1}, S_{j-1}) \text{ is matching} \\ e[i + 1, j - 1] & \text{otherwise} \end{cases}$$

Min total energy for the whole string is located at $e[1, n]$.

For any substring with a size greater than or equal to four, all possible bifurcations are considered as candidates for minimizing the total energy. Additionally, if the current substring's first and last nucleotides can match, then the pairing of those two is also considered. Out of the possible candidates, the minimum is taken and held by the corresponding position at the e matrix.

In addition, the b matrix is used to keep the minimum total energy of each substring with the assumption that its beginning and ending are paired. With this assumption, it is possible to recursively explore the stacks without making wrong assumptions about whether certain nucleotides are matched or not. If one inner level of pair is also matching along with the current pair, the energy of the formed stacking pair is added on top of the minimum energy obtained

from the inner substring. Otherwise, the inner pairing cannot be made, and so, the energy obtained from the inner substring is returned directly.

Results:

Command line output:

Pairs: [(1, 34), (2, 33), (3, 26), (4, 25), (5, 24), (6, 23), (14, 22), (15, 21), (16, 20), (40, 69), (41, 68), (44, 61), (45, 60), (46, 59), (51, 58), (52, 57)]

Matches: [('G', 'C'), ('A', 'U'), ('U', 'A'), ('G', 'C'), ('G', 'C'), ('G', 'C'), ('C', 'G'), ('G', 'C'), ('A', 'U'), ('G', 'C'), ('G', 'C'), ('G', 'C'), ('G', 'C'), ('U', 'A'), ('A', 'U'), ('G', 'C')]

Stacks: [(((('G', 'C'), ('A', 'U'))), -2.2), (((('U', 'A'), ('G', 'C'))), -1.5), (((('G', 'C'), ('G', 'C'))), -3.4), (((('G', 'C'), ('G', 'C'))), -3.4), (((('C', 'G'), ('G', 'C'))), -3.3), (((('G', 'C'), ('A', 'U'))), -2.2), (((('G', 'C'), ('G', 'C'))), -3.4), (((('G', 'C'), ('G', 'C'))), -3.4), (((('G', 'C'), ('U', 'A'))), -1.5), (((('A', 'U'), ('G', 'C'))), -2.2)]

Total Energy: -26.5

Number of pairs: 16

Optimal value: -26.5

Execution time: 0.022002458572387695 seconds

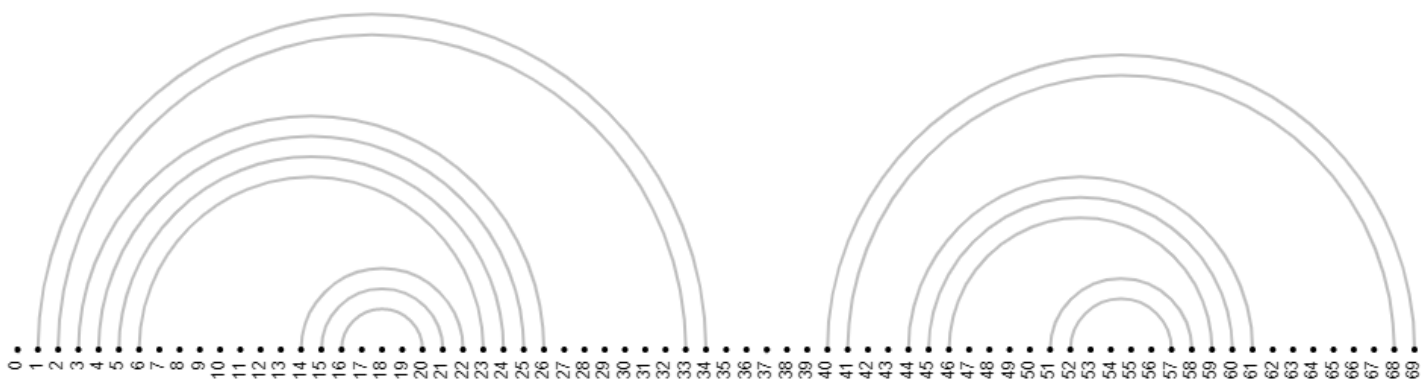


Figure 6: RNA matching diagram for Part F