Algorithm for Bioinformatics Exercise 2 David Riemer, Usama Mehmood

Problem 1

Our team created the following Map:

Luciferase in nature

What are Reporter Genes?

"Genetic tools used in molecular biology to easily track and measure gene expression"

=tiny "light bulbs" inside cells that glow or shine when a gene turns on

How?

1. Insert reporter gene downstream of a promoter

2. If the promoter is active -> reporter gene is expressed

3. The product (fluorescent or luminescent signal) indicates gene activity

Examples of Reporter Genes

GFP (Green Fluorescent Protein) emits green fluorescence when excited by blue light:
It makes tumors, Alzheimer's plaques, and even pathogenic bacteria visible

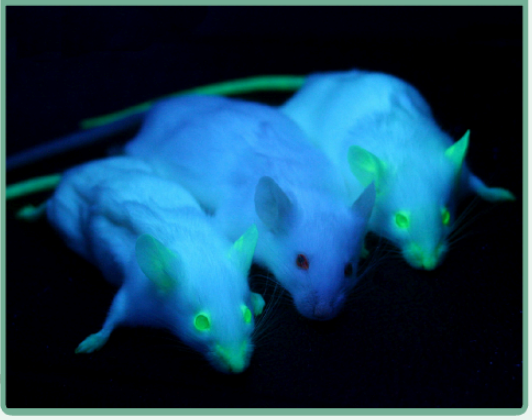
Luciferase catalyzes the oxidation of luciferin, producing bioluminescence proportional to gene activation

Reporter systems revolutionized molecular biology by turning invisible gene expression into observable, quantifiable signals!!!

They are used in cancer biology, stem cell tracking, drug screening, transfection validation, and synthetic biology circuits

More Applications!

On the left and right, a GFP mouse under UV light.



GFP in nature

Luciferase in nature

David R, Usama M

Sources:
https://www.promega.de/en/resources/guides/cell-biology/bioluminescent-reporters/
https://de.slideshare.net/slideshow/reporter-genes-249800990/249800990
https://www.quora.com/Whats-a-reporter-gene-and-whats-its-importance#:~:text=A%20reporter%20gene%20is%20used,1
CHALFIE, Martin. GFP: Ein Protein bringt Licht ins Dunkel (Nobel-Vortrag). Angewandte Chemie, 2009, 121. Jg., Nr. 31, S. 5711-5720.
CHALFIE, Martin, et al. Green fluorescent protein as a marker for gene expression. Science, 1994, 263. Jg., Nr. 5148, S. 802-805.

The current version can be found at
https://app.infinitymaps.io/maps/qM9HDL2BDjm/p7j76QD87jF

Problem 2

3-Partition-Problem

The takes a total of $3n$ integers, calles $X = x_1, ...x_{3_n}$ and splits them in n triplets with $x_{j_1} + x_{j_2} = x_{j_3}, j = 1, ..., n$, in which case the triplets cover all numbers and are all disjoint.

Inexact-Turnpike-Problem

The ITP takes a list of $\binom{n}{2}$ intervals $[d_i, d_i], i = 1, ..., \binom{n}{2}$, so instead of a multiset input, the problem uses distances and it searches, wether or not there is a sequence of prefix sums such that every entry in the difference multiset of that sequence maps to exactly one distcance.

To show that the Inexact-Turnpike-Problem (ITP) is NP-hard, we use the already known NP-hard problem "A + B = C"-3-Partition (3PP) as a starting point. Each instance of the 3PP is transformed in such a way that the resulting numbers represent the distances of a Inexact-Turnpike instance.

$$“A + B = C” - 3 - Partition \leq_P Inexact - Turnpike - Problem$$

Reduction

To start the reconstruction of the ITP-Instance, we first set some variables.

$$B = max(X)$$
$$M = 10B$$

We set B as the largest number in the 3PP-input. we then set M as a large number, in this case 10 times B.

We then construct the $\binom{N}{2}$intervals for our ITP as follows:

For every $x \in X$ wie create exactly one interval $[x, x]$, which are $3n$ intervals, representative of every number.

We want to create these small "islands" of 3 points and force these blocks to be far apart from each other. To build the remaining $\binom{3n}{2} - 3n$ intervals responsible for the distance between these islands we choose to build intervals which are multiples of M.

$$[kM - B, kM + B], k \in \{1, ..., n - 1\}$$

The distances between points of 2 different groups or islands are always multiples of M. The k gets larger as the distance between the different "islands" increases, with the biggest amount of k representing the outer most islands.

To prove that if the 3PP exists, a solution for the ITP also exists, we assume that an X exists, so that X can be split into triplets of form (a_j, b_j, c_j), with characteristic $a_j + b_j = c_j$.

We then create n blocks with $A_j = j * M$ representing the first number of each block. In each block of 3 numbers, we set those points as $S = \{A_j, A_j + a_j, A_j + c_j\}$.

For example we set $A_0 = \{0, 3, 10\}$, then $a_0 = 3, c_0 = 10$ and $b_0 = 7$, resulting in the distances within one of the blocks (these points also satisfy $3 + 7 = 10$). These points are the $3n$ small intervals we set at the beginning.

All the intervals between the blocks are of form $|A_{j_1} - A_{j_2}| = |(j_1 - j_2)M|$, which all lay inside the defined large intervals.

The other direction of the proof is that if the ITP-instance is solvable, a partitioning into triples also exists.

When asking ChatGPT how to approach this proof, it just assumes that there is a solution of points S and that this S splits blocks of 3 elements each, however we could not get a rigoros proof for how this is done. We will just continue the proof with S and that characteristic.

Every interval [x, x] with $x \leq B$ must correspond to a distance x. A distance between different blocks would be at least $M - B > B$. Therefore, all $3n$ small intervals must be realized within the same block.

We then assume that S can be split in n small triplets with $p < q < r$, in which the distances of the 3 points are all $\leq B$, and the 2 smalelr distances being able to be added up to the largest distance (between p and r).