

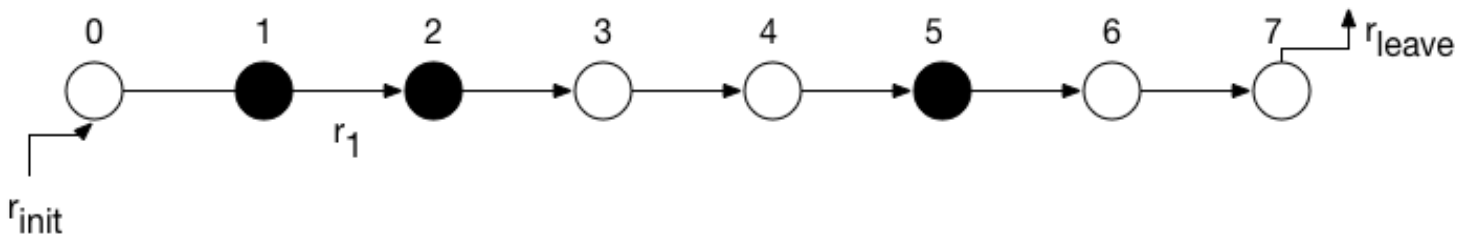
Lecture 15: Programming Example: TASEP

The goal for this lecture is to implement a reasonably large program from scratch. The task we will program is to simulate ribosomes moving along a mRNA transcript during the process of translation.

The TASEP model

The TASEP model is a model used to simulate systems such as traffic, queues, and ribosomes. The model uses a 1-dimensional grid of a fixed, finite length L . Particles (such as cars) enter at the left end of the grid. They move, at rates that depend on their location, to the right until they fall off the grid.

Only one particle can occupy any grid position at any time. A particle cannot move forward until the space it would move into is empty. For example:



Here, the solid circles represent positions occupied by items. The item at position 1 can't move until the item at position 2 moves.

TASEP stands for the features of this model:

- Totally Asymmetric: particles can only move forward
- Simple Exclusion: only one particle can be at each grid point
- Process: it describes dynamics

Translation

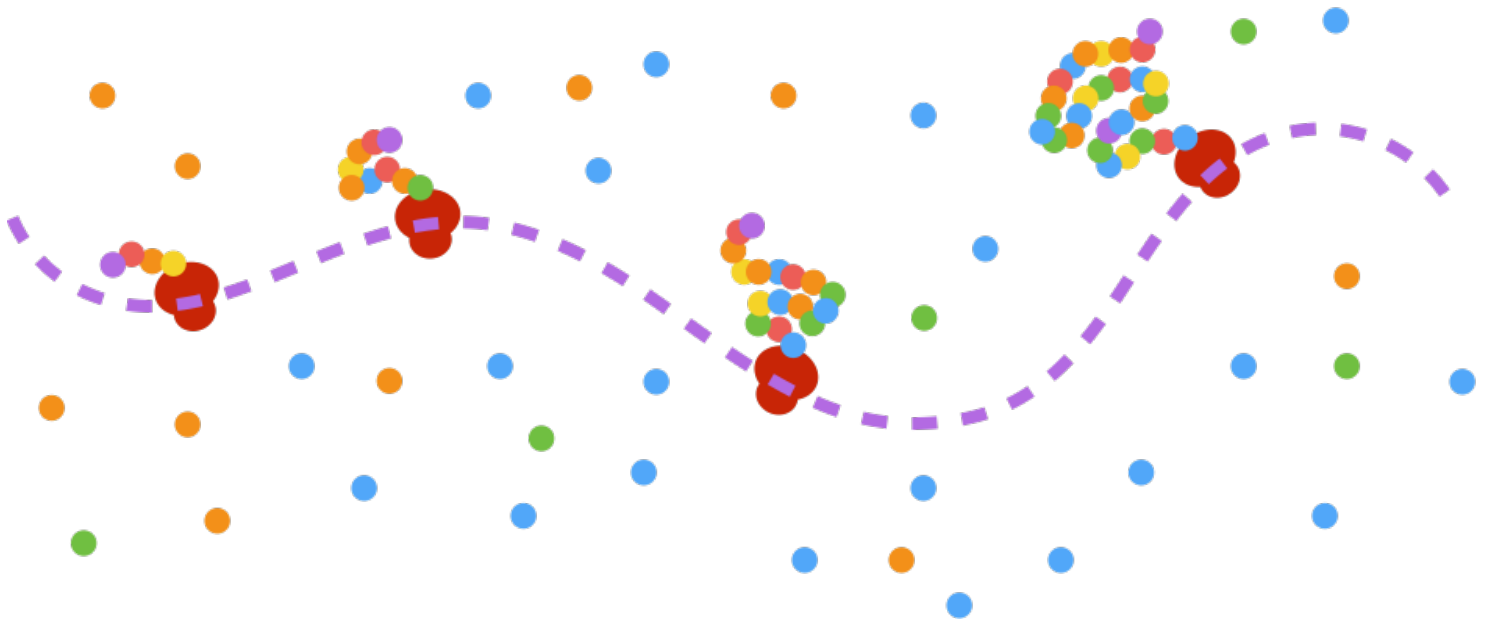
DNA encodes for genes, which are subregions of the DNA molecule. These subregions are transcribed into messenger RNA (mRNA) molecules during a process called transcription which is the first step toward making a protein corresponding to the gene.

Following transcription, molecular machines called ribosomes walk down the mRNA molecule translating it into a protein. The genetic code is a mapping between three-letter-long "words" of RNA called codons to amino acids.

A ribosome binds to the mRNA at the start of the transcript, moves down the transcript one codon at a time (and never backs up). The speed at which the ribosome moves is likely somewhat dependent on the type of codon it is translating. At the end of the transcript ribosomes fall off.

A ribosome binding to the start of the transcript is called initiation.

Many ribosomes can be translating a mRNA molecule at the same time, but no two ribosomes can occupy the same place on the mRNA. The picture is this:



where the red blobs are the ribosomes translating the mRNA (purple dashed line). The chains of circles attached to the ribosomes are the proteins being created.

Translation looks a lot like the idealized TASEP model. We will simulate ribosome motion using this model.

The simulator

We will simulate for a given number of time steps t .

Input data:

The codon-specific rates will be given in a file formatted as below:

1	init	200
2	TTT	55
3	TTC	110
4	TTA	70
5	TTG	135
6	TCT	110
7	TCC	55

The first column gives the codon, the second column gives the number of time steps that a ribosome must wait at a given codon. The special "codon" `init` is the time between initiation events.

Denote by `rate["TTT"]` the delay for codon type `TTT`. A ribosome must wait `rate[x]` timesteps before moving if it is sitting at a codon of type `x`.

The sequence of the mRNA will be provided in a standard format called FASTA, which looks like this:

```

1 >YAL001C TFC3 SGDID:S000000001, Chr I from 151006-147594,151166-151097, Genome Release 64
2 ATGGTACTGACGATTTATCCTGACGAACTCGTACAAATAGTGTCTGATAAAATTGCTTCA
3 AATAAGGGAAAAATCACTTTGAATCAGCTGTGGGATATATCTGGTAAATATTTTGATTTG
4 TCTGATAAAAAAGTTAAACAGTTCGTGCTTTCATGCGTGATATTGAAAAAGGACATTGAG
5 GTGTATTGTGATGGTGCTATAACAACAAAAATGTGACTGATATTATAGGCGACGCTAAT
6 CATTCACTCGGTTGGGATTACTGAGGACAGCCTATGGACATTATTAACGGGATACACA
7 AAAAAGGAGTCAACTATTGGAAATCTGCATTTGAACTACTTCTCGAAGTTGCCAAATCA
8 GGAGAAAAAGGGATCAATACTATGGATTTGGCGCAGGTAAGTGGGCAAGATCCTAGAAGT

```

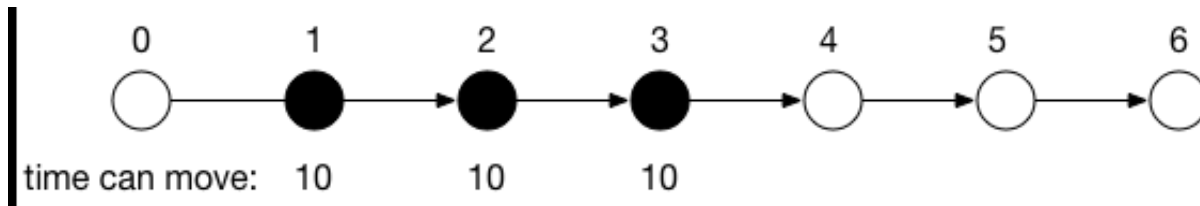
Output:

We would like to know a profile of where ribosomes were sitting over the course of the simulation. To that end, we will produce output that looks like that:

1	0	ATG	998360	(110)
2	1	GTA	997685	(65)
3	2	CTG	997220	(15)
4	3	ACG	997675	(35)
5	4	ATT	998925	(135)
6	5	TAT	998000	(40)
7	6	CCT	998085	(20)
8	7	GAC	999140	(160)
9	8	GAA	998970	(150)
10	9	CTC	998390	(10)
11	10	GTA	998790	(65)
12	11	CAA	998885	(90)

What happens at a timestep of the simulation:

At each time step, all the ribosomes that can move should move. A ribosome at codon-type `x` can move if it's waited `rate[x]` timesteps and if there is no ribosome at the next codon. If there is a chain of ribosomes that are ready to move at a timestep, then they all should move:



In the above example, the ribosomes at positions 1, 2, and 3 should all move at time 10 one space to the right.

Strategy for the Code

We will follow the following steps to write the code:

1. Write code to read in the sequence and the rates --- starting with reading your data is a good way to make early progress.
2. We will then decide on a data model. Something like:

```

1 | type Transcript []bool // t[i] == true if posn i has ribosome
2 |
3 | type Ribosome struct {
4 |     posn int
5 |     timeCanMove int
6 | }

```

3. We will then switch to a top-down design approach where we write the outer loop of the simulation. Something like:

```

1 | for curTime := 1; curTime < maxTime; curTime++ {
2 |     // find a ribosomes to move
3 |     // move them
4 |     // initiate a ribosome if we can
5 |     // update our profile of where the ribosomes are
6 | }

```

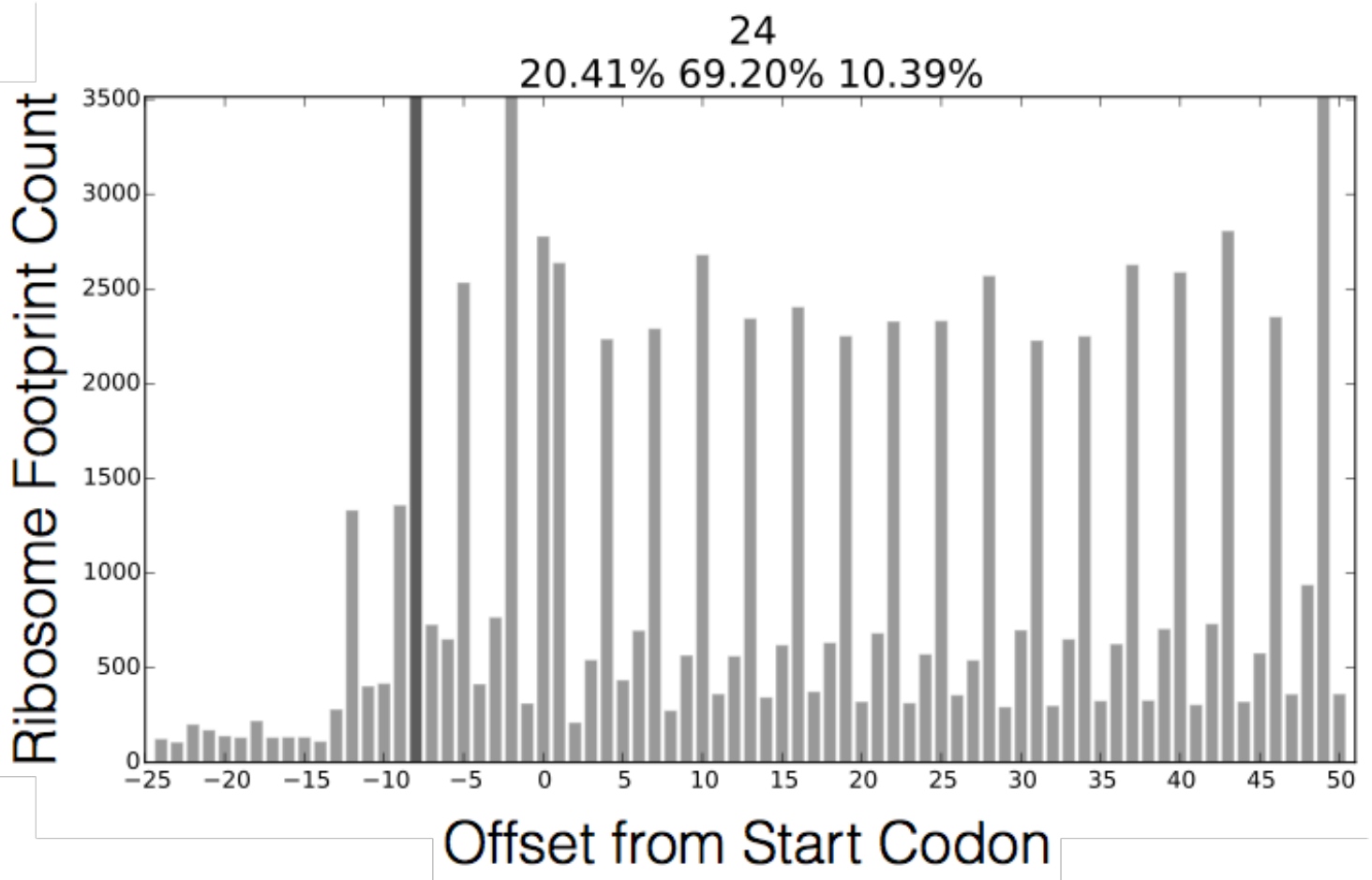
The code we came up with is available separately.

Ribo-seq

The experimental protocol called ribo-seq measures the positions of ribosomes that are translating mRNA. It works by:

1. Attaching the ribosomes where they are onto the mRNA
2. Degrading all the mRNA that is not protected by a ribosome
3. Removing the ribosomes from the mRNA they are stuck to
4. Sequencing the mRNA that was protected by the ribosomes. This gives you sequences of the the "ribosome footprints"
5. Mapping these sequences to the know sequences of the genes

The above approach produces profiles of ribosome locations:



This profile is similar to that produced by our simulation (except that the above figure shows that sometimes ribosomes are captured in between codons --- the large peaks are the codon peaks and the smaller peaks are ribosomes that look like they were in between codons when they were frozen).

Exercises

Extend our simulator to:

- draw a graph of ribosome positions
- to not waste time "simulating" for timesteps in which nothing moves
- automatically find rates that match an input profile (hard)

Summary

- Translation can be simulated using the TASEP model
- This is a moderately sophisticated program (that needn't use anything we haven't already seen in depth)

- How fast ribosomes move and their dynamics is an active research area

Glossary

- transcription: the process of producing an mRNA template from the genome.
- translation: the process of constructing a protein from an mRNA template.
- ribosome: a molecular machine that reads an mRNA template and produces a protein from it
- initiation: the process by which a ribosome bind to an mRNA and starts translating
- ribo-seq: experimental technique to measure ribosome locations
- codon: a 3-letter code for an amino acid
- genetic code: the complete set of mapping from 3-letter codes to amino acids