

Visualization of large RNA molecules

Problem:

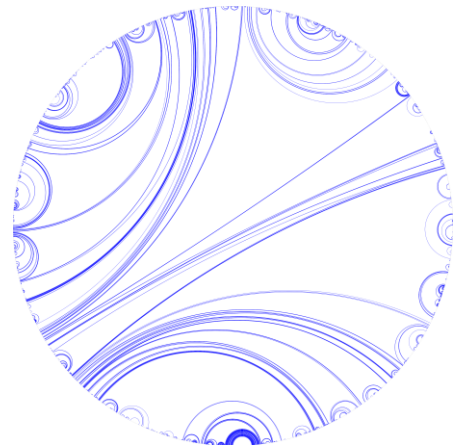
Visualization of data is often a problem in biology. An example is in RNA structure prediction: it is now possible to predict the secondary structure of big RNA molecules, but in order for biologists to be able to actually use these results, they need to be able to visualize them. One very useful way of representing large structures is the so-called “circular Feynmann” diagram. It is a very simple representation, but currently there are very few programs that can plot such figures at all, and none of them are fast and stable enough for very big RNAs, eg. the genomes of viruses. This project would therefore fulfill a great need in biology by creating a user-friendly program that can plot the secondary structure of very large RNA molecules.

Project description:

Basic requirements:

Create a program that can:

- read in structure data from a specifically formatted text file (the format is given)
- draw a circle and divide it into equal segments
- connect specific points on the circle’s circumference, on the basis of the input data (see figure to the right)
- add numbers at reasonable intervals to the circle’s circumference (indicating position in the data)
- export this figure in some kind of image format, preferably both .png and .eps



It is imperative that the program is:

- platform-independent (Win, Linux, Mac)
- very fast and very light: it needs to be designed to manipulate *very large* amounts of data in a reasonable time on a normal computer (thousands of connections in a circle with tens of thousands of possible connection points on its circumference should display without issues)
- stable and very user-friendly, very easy for a biologist to install, run and use

Ensuring that the performance requirements are met should be very high priority from the start.

Extensions (do none, any or all – listed here in rough order of increasing difficulty):

- carry out simple statistics on the input data on the basis of pre-defined formulae, and display the results in a nice way, with possibility for exporting them
- add possibility for plotting data from several files in the same figure, in different colors
- color lines common to data from all input files with a particular color
- add user-friendly controls, such as zoom, spacing, labels etc.
- recognize non-nested structures and color/emphasize them in a different way
- integrate all of it into the CLC Workbenches as a plugin
- add functionality for user interaction, especially suitable for interactive screens/tablets etc
- develop and implement different, novel ways of visualizing this kind of data (through consultation with biologists to identify their needs)