Dot Bracket Notation for Nucleic Acid Nanostructures

Slides by Reem Mokhtar

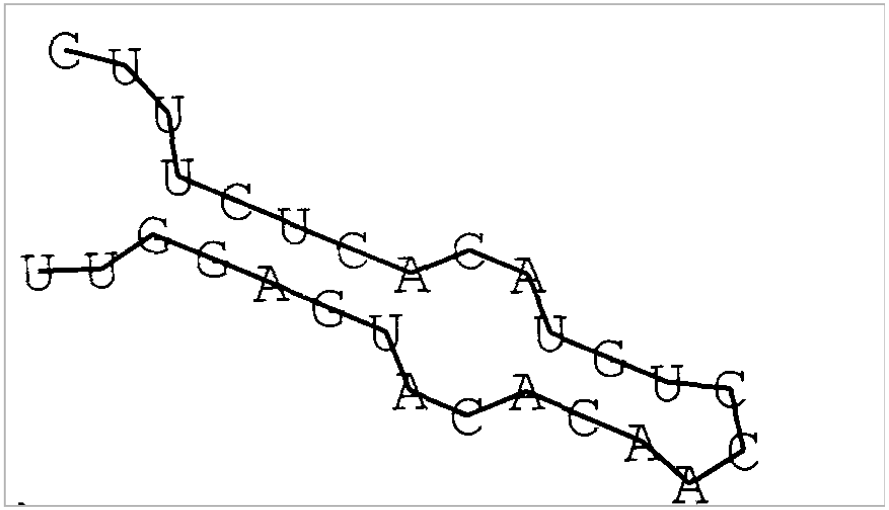
Dot Bracket Notation for RNA and DNA nanostructures

- Purpose:
  - Ease of interaction and design
  - Aid in validating designs
- Representations might include
  - GUI input
  - Rendered results
  - Back-end data structures

Dot Bracket Notation

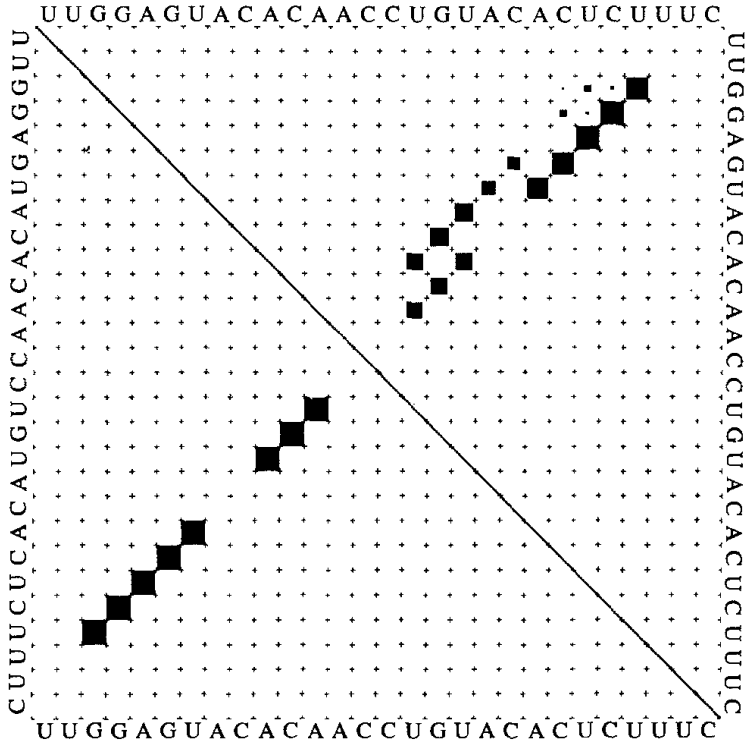
```
tram> RNAfold -T 42 -p1
Input string (upper or lower case); @ to quit
.........1.........2.........3.........4.........5.........6.........7 .........
UUGGAGUACACAACCUGUACACUCUUUC
length = 28

UUGGAGUACACAACCUGUACACUCUUUC
..(((((..(((...)))..)))))...
minimum free energy = -3.71
..(((((([[(,,...)))..)))))...
free energy of ensemble = -4.39
frequency of mfe structure in ensemble 0.337231
```



Hofacker, I. L., Fontana, W., Stadler, P. F., Bonhoeffer, L. S., Tacker, M., & Schuster, P. (1994). Fast folding and comparison of RNA secondary structures. Monatshefte für Chemie/Chemical Monthly, 125(2), 167–188.

Dot Bracket Notation

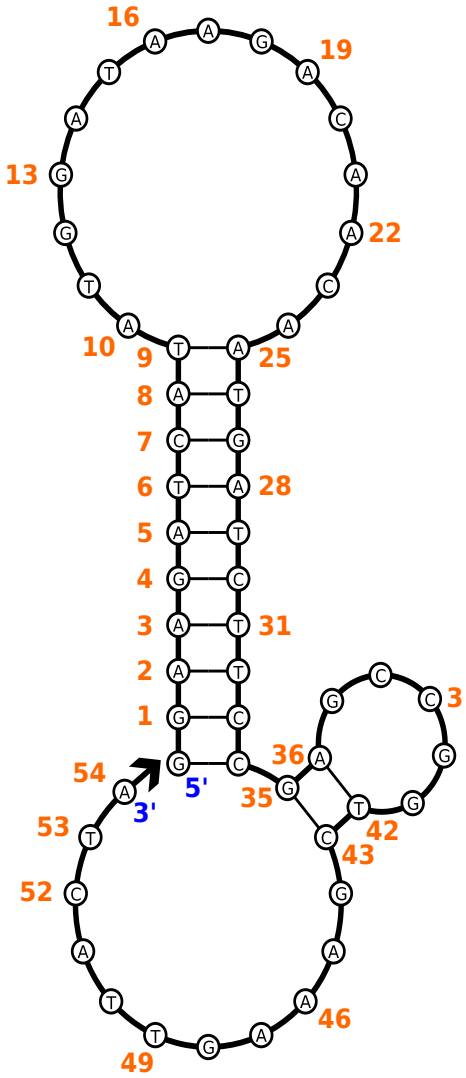


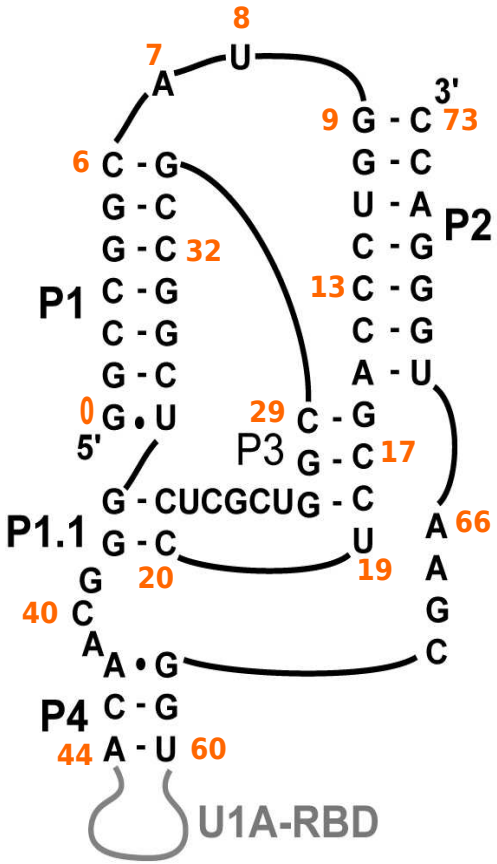
c

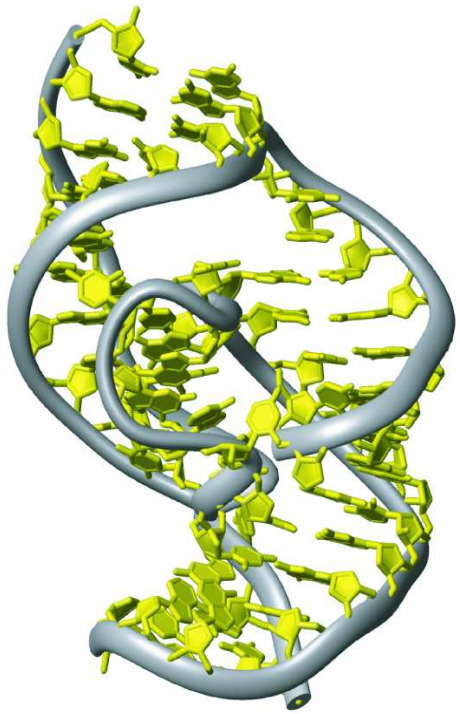
Fig. 1. Interactive example run of RNAfold for a random sequence. When the base pairing probability matrix is calculated by symbols . , | { } ( ) are used for bases that are essentially unpaired, weakly paired, strongly paired without preferred direction, weakly upstream (downstream) paired, and strongly upstream (downstream) paired, respectively. Apart from the console output, **a**, the two postscript files rna.ps, **b**, and dot.ps, **c**, are created. The lower left part of dot.ps shows the minimum energy structure, while the upper right shows the pair probabilities. The area of the squares is proportional to the binding probability

Hofacker, I. L., Fontana, W., Stadler, P. F., Bonhoeffer, L. S., Tacker, M., & Schuster, P. (1994). Fast folding and comparison of RNA secondary structures. Monatshefte für Chemie/Chemical Monthly, 125(2), 167–188.

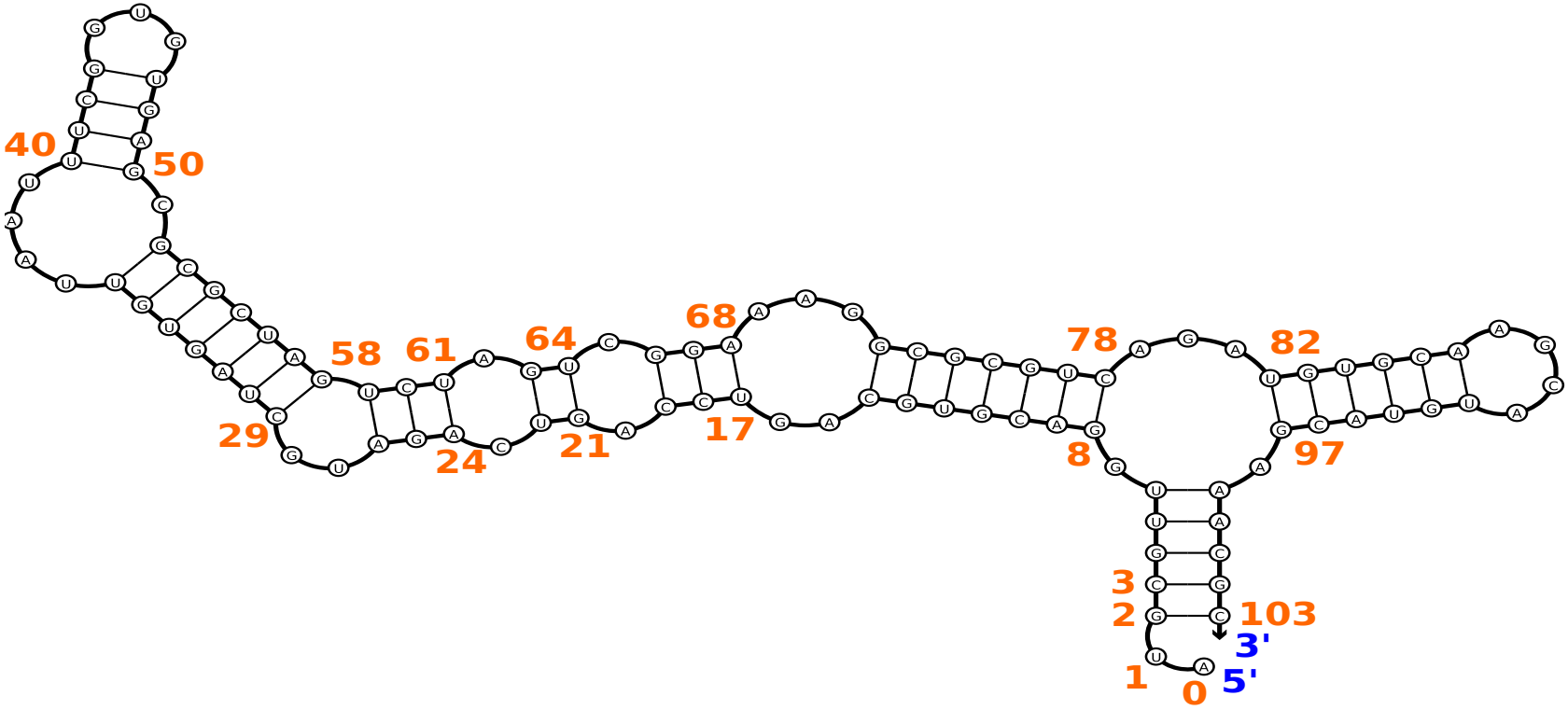
Another Example of a Nucleic Acid Nanostructure that can be specified by Dot-parenthesis Notation



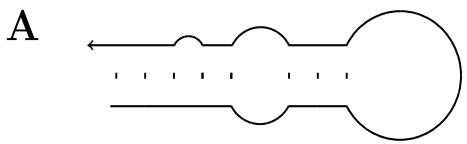




Another Example of a Nucleic Acid Nanostructure that can be specified by Dot-parenthesis Notation



Extended Dot Bracket Notation



B CCGAUAGAGGCGUGCGGUCAAGGUCCGG
C ((((((...(((.....)))...)).)))

Figure 1: A sample secondary structure of an RNA molecule is shown (A) together with the notation of its sequence (B) and its structure in dot-bracket form (C).

Matches parenthesis and dots to denote paired and free bases, respectively.

Advantage: string denoting the secondary structure of a nucleic acid is of the same length as the string denoting the nucleotide sequence with a single character for each nucleotide.

Disadvantages:

- the secondary structure is of importance, but the detailed sequence that yields the structure is over large stretches arbitrary.
- In such a scenario the dot-bracket notation is often cumbersome, it leads to large expressions with information that is partially obscured for human readers, because it would require counting identical characters.
- While the sequence is typically arbitrary in most positions, as long as the structure of the molecule is preserved, nucleic acids with functional properties, such as catalytic activity, often require specific bases in a few positions.
- If in a few places the nucleotide sequence (i.e. the primary structure) is given, two strings are required: one to specify the structure in dot-bracket-notation and a second string to represent the type of the immutable nucleotides.
- Furthermore, for communicating structural features among humans a two-dimensional rendering of the one-dimensional dot- bracket notation is often desirable. It would be convenient if specific features in the sequence could be communicated to the rendering software.

Extended Dot Bracket Notation

	Description	Usage	Comment
`[ ]`	Grouping of base positions	`[.8]@{label A}`	Eight unbound bases marked as “label A”.
`{ }`	Parameter delimiter		see example above
`{ }`	Set delimiter	`.:{A,C}`	A single unbound base that can be either A or C.
`{ }`	Repetition delimiter	`A{10}`	Always optional.
`_`	Line width	`((([.5]_1)))`	Stem-loop structure with bold loop
`$`	Colour	`(3[.2]$1(3.4)3)3`	A buldge in red.
`~`	Line decoration	`.24~1(3.3)3`	Binding site marked as crinkled line.
`@`	Annotation marker		See first row.
`+`	Multi-molecule binding	`(24+1(3.3)3`	Sticky end of 24 bases, will bind to site marked 1 on other molecule.
`:`	Base assignment	`)):A`	Two binding bases, the second one of which is A; See also set delimiter.
`^`	Base exclusion	`(((..^U.)))`	Stem loop where the central base in the loop is not a uracil.
`%`	Clevage point	`(((..%..(((`	Between bases, i.e., not a base position.