



Accepted Data Formats

Bpseq format

[Input and Output]

The structural information in the bpseq format is denoted in three columns. The first column contains the sequence position, starting at one. The second column contains the base in one-letter notation. The third column contains the pairing partner of the base if the base is paired. If the base is unpaired, the third column is zero.

The bpseq format is used on the [Comparative RNA Web \(CRW\)](http://www.rna.ccbb.utexas.edu). The files from the CRW Site contain four header lines, describing the filename, organism, accession, and a citation. These header lines may be included in the data. The parser recognizes this specific header format, and will not accept arbitrary header lines. The parser expects **one** sequence/structure record per file.

Without header

Including header

```

1 C 0      Filename: test.bpseq
2 C 9      Organism: Some organism
3 U 8      Accession Number: XYZ123
4 G 0      Citation and related information available at http://www.rna.ccbb.utexas.edu
5 A 0      1 C 0
6 A 0      2 C 9
7 C 0      3 U 8
8 A 3      4 G 0
9 G 2      5 A 0
           6 A 0
           7 C 0
           8 A 3
           9 G 2

```

Connect (.ct) format

[Input and Output]

The connect format is column based. The first column specified the sequence index, starting at one. Columns 3, 4, and 6 redundantly give sequence indices (plus/minus one). The second column contains the base in one-letter notation. Column 4 specifies the pairing partner of this base if it involved in a base pair. If the base is unpaired, this column is zero.

The parser expects **one** header line containing the word "ENERGY", "Energy", or "dG". Arbitrary header lines will not be accepted. Files in connect format may contain multiple sequence/structure records. The specified pseudoknot removal method will be applied to all records in the file.

```

73 ENERGY =      -17.50      S.cerevisiae_tRNA-PHE
1 G      0      2 72      1
2 C      1      3 71      2
3 G      2      4 70      3
4 G      3      5 69      4
5 A      4      6 68      5
6 U      5      7 67      6
7 U      6      8 66      7
8 U      7      9  0      8

.
.
.

66 A      65      67      7      66
67 A      66      68      6      67

```

68 U	67	69	5	68
69 U	68	70	4	69
70 C	69	71	3	70
71 G	70	72	2	71
72 C	71	73	1	72
73 A	72	74	0	73

(adapted from <http://www.binf.ku.dk/~pgardner/bralibase/RNAformats.html>).

Vienna format

[Output only]

The Vienna format or dot-bracket format is a string notation for a **nested** RNA structure. An unpaired base is denoted with a dot, a base pair is denoted with an opening and closing bracket. The opening bracket corresponds to the upstream partner, the closing bracket to the downstream partner. The Vienna format cannot denote pseudoknots and is therefore only available as output format.

We return a Vienna structure in three lines: a header line starting with a > sign, the RNA sequence on a single line, and the Vienna string on a single line.

```
>Header line
AUCGAGAAUCGAAC
..(((....)))...
```

Additional formats

Please contact Sandra Smit (S.Smit at few.vu.nl) to request the support of other RNA structure formats or if you experience problems getting your data accepted by the parsers.

When using our method, please cite:

Sandra Smit, Kristian Rother, Jaap Heringa, and Rob Knight. *From knotted to nested RNA structures: a variety of computational methods for pseudoknot removal*. (2008) RNA 14(3): 410-416.

Copyright © 2007 Sandra Smit, Kristian Rother, Jaap Heringa, and Rob Knight