Package 'RRNA'

July 10, 2024

Type Package

Title Secondary Structure Plotting for RNA

Version 1.2

Date 2024-07-10

Description Functions for creating and manipulating RNA secondary structure plots.

License GPL-3

NeedsCompilation no

Author JP Bida [aut],
Jonathan Price [cre, ctb] (<https://orcid.org/0000-0001-6554-5667>)

Maintainer Jonathan Price <jlp76@cam.ac.uk>

Repository CRAN

Date/Publication 2024-07-10 10:20:02 UTC

Contents

RRNA-package *RNA secondary structure ploting*

Description

Set of functions for creating and manipulating RNA secondary structure plots from CT files or bracket notations.

Details

Package:	RRNA
Type:	Package
Version:	1.0
Date:	2015-07-27
License:	GPL-3

Author(s)

JP Bida Maintainer: JP Bida <bida.john@gmail.com

Examples

```
### Create a CT file from bracket notation
ct=makeCt("(((...(((...)))...(((...)))...)))","AAAUUUCCCAAAGGGUUUAAAGGGUUUCCCUUU")
coord=ct2coord(ct)
RNAPlot(coord,hl=c("GGGUUU","AAAUUU"),seqcols=c(2,4),labTF=TRUE)
```

alignCoord *Alignment of secondary structure folds to 2 nucleotides.*

Description

Given a coordinate file with multiple RNA secondary structures, it aligns all folds such that n1 is at position (x,y) and n2 has its y coordinate equal to y

Usage

```
alignCoord(data, n1, n2, x, y)
```

Arguments

data	R data frame containing the coordinates for plotting a given secondary structure
n1	Nucleotide position that will be translated to (x,y)
n2	Nucleotide position that will have its y coordinate equal to y
x	x coordinate that n1 will be translated to
y	y coordinate that n1 will be translated to

Value

Returns a data frame containing fold coordinates.

Author(s)

JP Bida

See Also

RNAPlot

Examples

```
### Create two RNA secondary structures ####
ct1=makeCt(
           "((((...(((((((....)))))))...((((...))))...))))",
           "CCCCAAAGGGGGGGAUUACCCCUCCUUUAAAAGGGUUUUCCCCCCC"
          )
ct2=makeCt(
           "((((...(((((((....)))))))..((..((...))))...))))",
           "CCCCAAAGGGGGGGAUUACCCCUCCUUUAAAAGGGUUUUCCCCCCC"
          )

### Create a coordinate file ####
dat1=ct2coord(ct1)

### Each RNA fold needs its own id ###
```

```
dat1$id=1

#### Create a coordinate file ####
dat2=ct2coord(ct2)
### Each RNA fold needs its own id ###
dat2$id=2

dat=rbind(dat1,dat2)

adat=alignCoord(dat,1,46,0,0)

### Plot the aligned RNA folds ####
RNAPlot(adat[adat$id==1,])
l=length(adat$seq[adat$id==2])
RNAPlot(adat[adat$id==2,],modspec=TRUE,modp=c(1:l),modcol=rep(4,l),mod=rep(16,l),add=TRUE)
```

`aptPlotCT` *RNA secondary structure plotting from CT files*

Description

Generates and RNA secondary structure plot from a CT file. Removes pseudoKnots automatically and allows them to be drawn back in with pseudoTF=TRUE.

Usage

```
aptPlotCT(file, ranges = 0, add = FALSE, hl = NULL, seqcols = NULL,
          seqTF = FALSE, labTF = FALSE, nt = FALSE, dp = 0.5,
          modspec = FALSE, modp = NULL, mod = NULL, modcol = NULL,
          tsize = 0.5, main = "", pseudoTF = FALSE, pseudo_nums = NULL,
          ticks = NULL, ticksTF = FALSE
         )
```

Arguments

`file`	CT file name
`ranges`	A data frame containing the ranges of sequence positions that should be highlighted with given colors. `ranges=data.frame(min=c(69,1,7),max=c(74,5,17),col=c(2,3,4),des 1","Region 2","Region 3"))` The above will highlight the nucleotides at positions 69-74, 1-5, and 7-17 respectively
`add`	Should the new plot be added to an existing plot TRUE/FALSE
`hl`	Takes an array of sequences and highlights them with seqcol `hl=c("GGGAAAA","GGGCCCC")` The above hl will highlight the nucleotides in the secondary structure that have the given sequences with the colors provided in the seqcols option.
`seqcols`	Colors that should be used to highlight the sequences given in hl
`seqTF`	If sequence is a vector set as TRUE

labTF	TRUE/FALSE plot the legend
nt	TRUE/FALSE plot the nucleotide sequence on the secondary structure
dp	Floating point value to determine how far from the coordinates the nucleotide sequence should be plotted. Values between 0 and 5 usually work best.
modspec	TRUE/FALSE modify specific positions in the secondary structure. Used in combination with modp,mod,and modcol. This allows you to change the shape and color of nucleotide in the secondary structure.
modp	Array defining the specific positions to be modified in the plot `modp=c(1:10)`
mod	Array defining the pch values to be plotted at the positions given by modp. `mod=c(rep(15,5),rep(16,5))`
modcol	Array of color values to be used for plotting at the positions defined by modp in the secondary structure. `modcol=rep(4,10)`
tsize	Text size used for plotting the nucleotide sequence in the secondary structure. Only applicable when nt=TRUE. Values between 0.1 and 4 work well.
main	Title used for the plot when labTF is set to TRUE.
pseudoTF	Plot pseudo knot sequences
pseudo_nums	indices of the nucleotides included in pseudoknots
ticksTF	TRUE/FALSE include ticks
ticks	Positions where the ticks should be drawn. These are sequence positions in the RNA molecule

Value

Returns and R plot object

Author(s)

JP Bida

See Also

RNAPlot

Examples

```
### PseudoKnots ###
pk= makeCt("((((...(((((((.........)))))))...((((.........))))...))))",
          "AAAAAAAACCCCCCCCAAAGGGGGGGAUUACCCCUCCUUUAAAAGGGUUUUCCCCCCC"
         )
pk$bound[pk$pos==20]=42
pk$bound[pk$pos==19]=43
pk$bound[pk$pos==43]=19
pk$bound[pk$pos==42]=20
f = tempfile()
### Create a CT file for testing ###
write.table(pk[,c(1,4,2,3,6,5)],file=f,row.names=FALSE,col.names=TRUE)
```

```
aptPlotCT(f,ticksTF=TRUE,ticks=seq(1,60,by=5),pseudoTF=TRUE,pseudo_nums=c(19,20,43,42))
```

---

backward *Internal function for moving through secondary structures*

Description

Given a bracket notation for RNA secondary structure and an index of a ")" bracket type the backward function will find the "(" bracket that closes the ")" at the given index.

Usage

```
backward(stc, i)
```

Arguments

stc	Array of brackets and dots. `a=unlist(strsplit("(((...)))...((..))",""))`
i	index giving the position of a bracket

Value

returns the index of the bracket that closes the bracket at the given index

Author(s)

JP Bida

Examples

```
a=unlist(strsplit("(((...)))...((..))",""))
ind=backward(a,7)
```

---

bplfile *Creates a bpl file from a coordinate file*

Description

A bpl file can be created from a given coordinate file for inputing into other RNA visulatization programs

Usage

```
bplfile(dat, name)
```

Arguments

dat	Coordinate file created by ct2coord or loadCoords functions `ct=makeCt("((((...))))","AAAACCCUUUU ### Create the coordinate file ### dat=ct2coord(ct)`
name	Name of the file outputed

Value

Creates the file with the given "name"

Author(s)

JP Bida

Examples

```
ct=makeCt("((((...))))","AAAACCCUUUU")
### Create the coordinate file ###
dat=ct2coord(ct)

bplfile(dat,tempfile())
```

---

circleCoord *Internal function for finding the coordinates of NT's in a circle*

Description

Given an integer N the function returns N (x,y) coordinates for a polygon with N sides each of length 1. This is used to plot the loops in an RNA structure

Usage

```
circleCoord(n)
```

Arguments

n	Integer determining the number of sides

Value

Data frame with columns x,y defining coordinates of the polygons

Author(s)

JP Bida

Examples

```
pts=circleCoord(10)
plot(pts$x,pts$y)
```

ct2coord *Generate coordinate file*

Description

Creates a coordinate file from a CT file that has been loaded into a data frame

Usage

```
ct2coord(input)
```

Arguments

input	Data frame representing a ct file. Created from makeCt or loadCt

Value

Returns a coordinate file for the secondary structure represented in the CT file

Note

Pseudoknots sometimes cause trouble

Author(s)

JP Bida

See Also

RNAPlot

Examples

```
ct=makeCt("((((...(((((((....)))))))...((((...))))...))))",
          "CCCCAAAGGGGGGGAUUACCCCUCCUUUAAAAGGGUUUUCCCCCCC"
         )
coord=ct2coord(ct)
RNAPlot(coord)
```

ct2knet *creates a knet file from a CT file*

Description

Knet files are used as inputs for KnetFold secondary structure prediction program

Usage

```
ct2knet(file, ind = 0)
```

Arguments

file	Name of the CT file being converted to KnetFold file
ind	Index used to relabel sequence indexes

Value

Retuns a string containing the contains of the knet file

Author(s)

JP Bida

Examples

```
pk=makeCt("((((...(((((((.........)))))))...((((.........))))...))))",
          "AAAAAAAACCCCCCCCAAAGGGGGGGAUUACCCCUCCUUUAAAAGGGUUUUCCCCCCC"
         )
pk$bound[pk$pos==20]=42
pk$bound[pk$pos==19]=43
pk$bound[pk$pos==43]=19
pk$bound[pk$pos==42]=20

f = tempfile()
### Create a CT file for testing ###
write.table(pk[,c(1,4,2,3,6,5)],file=f,row.names=FALSE,col.names=TRUE)

### Convert CT file to Knet ###
out=ct2knet(f,0)
```

forward *Internal function for moving through secondary structures*

Description

Given a bracket notation for RNA secondary structure and an index of a "(" bracket type the forward function will find the ")" bracket that closes the "(" at the given index.

Usage

```
forward(stc, i)
```

Arguments

stc	a=unlist(strsplit("(((...)))...((..))","""))
i	Interger index

Value

Integer index

Author(s)

JP Bida

See Also

backward

Examples

```
a=unlist(strsplit("(((...)))...((..))",""))
ind=forward(a,1)
```

genCords *Internal function that generates coordinates for a given loop starting and stopping at p1 and p2 respectfully*

Description

Generates coordinates for a loop in a secondary structure. Internal function used by RNAPlot.

Usage

```
genCords(loop, p1, p2, input, vn)
```

Arguments

loop	List contianing a data frame that has the subset of nucleotides in a given loop
p1	The position of the first nucleotide in the loop
p2	The position of the second nucleotide in the loop
input	The data frame contianing the coordinate file for the entire RNA secondary structure
vn	A flag that flips over y axis if vn = 1.

Value

Returns a set of points

Author(s)

JP Bida

Examples

```
### This is an internal function ###
```

---

loadCoords *Loads a coordinate file into a data frame*

Description

Coordinate files can be created from the viennaRNA library.

Usage

```
loadCoords(filename)
```

Arguments

filename	Name of the coordinate file being loaded

Value

Data frame containing the coordinate file

Author(s)

JP Bida

References

The RRNAFold program generates the coordinate files used by RRNA

https://github.com/jpbida/ViennaScripts

Examples

```
### Create a test coordinate file using ct2coord ###
ct=makeCt("((((...(((((((.........)))))))...((((.........))))...))))",
          "AAAAAAAACCCCCCCCAAAGGGGGGGAUUACCCCUCCUUUAAAAGGGUUUUCCCCCCC"
         )
coord=ct2coord(ct)
### add an id ###
coord$id=1
f = tempfile()
### write out test file ###
write.table( coord[,c('id','x','y','seq','num','bound')],
             col.names=FALSE,row.names=FALSE,sep=",",file=f
           )

### Read in the coordinate file ##
input=loadCoords(f)

### Plot the file using RNAPlot ##
RNAPlot(input)
```

`loadCt` *Loads a CT file into an R data frame*

Description

A variety of RNA secondary structure prediction programs produce CT files. You can load these CT files into R using the loadCT function.

Usage

```
loadCt(file)
```

Arguments

`file`	The name of the CT file being loaded

Value

Returns at data frame containing the CT file data

Author(s)

JP Bida

See Also

`RNAPlot` `aptPlotCT`

Examples

```
### Create a CT file with PseudoKnots ###
pk=makeCt("((((...(((((((.........)))))))...((((.........))))...))))",
          "AAAAAAAACCCCCCCCAAAGGGGGGGAUUACCCCUCCUUUAAAAGGGUUUUCCCCCCC"
         )
pk$bound[pk$pos==20]=42
pk$bound[pk$pos==19]=43
pk$bound[pk$pos==43]=19
pk$bound[pk$pos==42]=20

### Create a CT file for testing ###

f = tempfile()
write.table(pk[,c(1,4,2,3,6,5)],file=f,
            row.names=FALSE,col.names=TRUE)

ctfile=loadCt(f)

### Before using ct2coord you need to remove the pseudo knots ###

l=pseudoKnot(ctfile)
dat=l[[2]]
cd=ct2coord(dat)
RNAPlot(cd)
```

---

loopLength *internal function that determines the length of a loop*

Description

Used by RNAPlot to get the length of a loop

Usage

```
loopLength(input, start)
```

Arguments

input	CT file `makeCt("((((...))))","AAAACCCUUUU")`
start	Position of the first nucleotide in the the loop

Value

Retuns a list contianing the output and stems

Author(s)

JP Bida

Examples

```
ct=makeCt("((((...((((..))))..((((...)))).))))","AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA")
out=loopLength(ct,4)
```

---

makeCt *make a CT file from a structure and sequence*

Description

Given an RNA secondary structure in bracket notation containing no pseudoKnots this function creates an R data frame that represents the secondary structures CT file.

Usage

```
makeCt(struct, seq)
```

Arguments

struct	Bracket notation. st="(((((....)))))..((..))"
seq	String containing the RNA sequence seq="AUAAUUAAAAAAAACCCCCAAA"

Value

Returns a data frame representing the bracket notaiton secondary structure in a CT file like format.

Author(s)

JP Bida

Examples

```
st="(((((....)))))..((..))"
seq="AUAAUUAAAAAAAACCCCCAAA"

ct=makeCt(st,seq)
```

pseudoKnot *removes pseudoknots from a ct file*

Description

internal function used to remove pseudoKnots before calling ct2coord

Usage

```
pseudoKnot(ctDat)
```

Arguments

ctDat	R data frame representing a CT file for RNA secondary structure

Value

Returns a list with the first item being a list of pseudoKnots and the second item being a CT file data frame with all pseudoKnots removed from the structure

Author(s)

JP Bida

See Also

RNAPlot, aptPlotCT, ct2coord,

Examples

```
pk=makeCt("((((...(((((((.........)))))))...((((.........))))...))))",
          "AAAAAAAACCCCCCCCAAAGGGGGGGAUUACCCCUCCUUUAAAAGGGUUUUCCCCCCC")
pk$bound[pk$pos==20]=42
pk$bound[pk$pos==19]=43
pk$bound[pk$pos==43]=19
pk$bound[pk$pos==42]=20

l=pseudoKnot(pk)

## Positions of removed pseudo knots ##
removed=l[[1]]

### clean ct file that can be used by ct2coord ###
ct=l[[2]]
```

RNAPlot *Generic RNA Secondary Structure Plotting Function*

Description

Given fold data from loadFolds or ct2coords RNAPlot plots the secondary structure

Usage

```
RNAPlot(data, ranges = 0, add = FALSE, hl = NULL, seqcols = NULL,
        seqTF = FALSE, labTF = FALSE, nt = FALSE, dp = 0.5,
        modspec = FALSE, modp = NULL, mod = NULL, modcol = NULL,
        tsize = 0.5, main = "", pointSize = 2, lineWd = 2)
```

Arguments

data	R data frame containing the coordinates for plotting a given secondary structure ### Example input file format: ### ### 0,158.534088,199.550888,G,0,-1 ### 0,152.741776,194.100571,A,1,-1 ### 0,149.307266,186.849899,A,2,-1 ### 0,148.749847,178.776566,G,3,-1 ### 0,151.196960,170.989944,C,4,59 ### 0,141.412643,159.620361,U,5,58 ### 0,131.628342,148.250793,U,6,57 ### 0,121.844025,136.881210,A,7,56 ### 0,112.059715,125.511642,C,8,55 ### 0,102.275398,114.142059,A,9,54 ### 0,89.142853,109.343330,A,10,-1 ### .... ### ### There is no header on the input file. The columns are ### ID,X,Y,SEQ,POS,BOUND ### ### ID - A unique ID for a given fold in the file ### X - X position of the NT in the secondary structure plot ### Y - Y position of the NT in the secondary structure plot ### SEQ - The nucleotide (A,G,U,C) ### POS - The position of the NT in the sequence ### BOUND - The position of the NT that the NT at POS is bound to
ranges	A data frame containing the ranges of sequence positions that should be highlighted with given colors. ranges=data.frame(min=c(69,1,7),max=c(74,5,17),col=c(2,3,4),des 1","Region 2","Region 3")) The above will highlight the nucleotides at positions 69-74, 1-5, and 7-17 respectively
add	Should the new plot be added to an existing plot TRUE/FALSE
hl	Takes an array of sequences and highlights them with seqcol hl=c("GGGAAAA","GGGCCCC") The above hl will highlight the nucleotides in the secondary structure that have the given sequences with the colors provided in the seqcols option.
seqcols	Colors that should be used to highlight the sequences given in hl
seqTF	If sequence is a vector set as TRUE
labTF	TRUE/FALSE plot the legend
nt	TRUE/FALSE plot the nucleotide sequence on the secondary structure
dp	Floating point value to determine how far from the coordinates the nucleotide sequence should be plotted. Values between 0 and 5 usually work best.

modspec	TRUE/FALSE modify specific positions in the secondary structure. Used in combination with modp,mod,and modcol. This allows you to change the shape and color of nucleotide in the secondary structure.
modp	Array defining the specific positions to be modified in the plot `modp=c(1:10)`
mod	Array defining the pch values to be plotted at the positions given by modp. `mod=c(rep(15,5),rep(16,5))`
modcol	Array of color values to be used for plotting at the positions defined by modp in the secondary structure. `modcol=rep(4,10)`
tsize	Text size used for plotting the nucleotide sequence in the secondary structure. Only applicable when nt=TRUE. Values between 0.1 and 4 work well.
main	Title used for the plot when labTF is set to TRUE.
pointSize	The size of points plotted in the secondary structure. Values betwen 0.1-5 work well.
lineWd	Line width for base pairings and backbone of secondary structures.

Value

Returns a generic R plot that can be used with the jpeg, postscript, etc. functions.

Author(s)

JP Bida

See Also

makeCt,loadCoords,ct2coord

Examples

```
## Create a CT file from bracket notation and sequence ###
ct=makeCt( "(((((...(((((((....)))))))...((((...))))...))))",
          "CCCCAAAGGGGGGGAUUACCCCUCCUUUAAAAGGGUUUUCCCCCCC"
)

## Create a coordinate file based on the CT file ###
dat=ct2coord(ct)

### Create a plot of the secondary structure ###
RNAPlot(dat)

### Plot positions 1:4 as green and 43:46 circles ##
### and show the legend
ranges=data.frame(min=c(1,43),max=c(4,46),col=c(2,3),
                  desc=c("Region 1","Region 2")
                 )
RNAPlot(dat,ranges,labTF=TRUE)

### Highlight the sequences CUCCU and CCCCAAA ###
```

```
RNAPlot(dat,hl=c("CUCCU","CCCCAAA"),seqcol=c(2,4),labTF=TRUE,main="RNA Molecule")

### Modify specific positions ####

RNAPlot( dat, modspec=TRUE, modp=c(1:4,43:46),mod=c(17,17,15,15,16,16,16,16),
         modcol=c(rep(2,2),rep(3,2),rep(4,4))
       )

### RNA Plot with nucleotides ###
RNAPlot(dat,nt=TRUE)

### RNA plot with nucleotides
RNAPlot( dat,nt=TRUE,modspec=TRUE,modp=c(1:4,43:46),
         mod=c(17,17,15,15,16,16,16,16),
         modcol=c(rep(2,2),rep(3,2),rep(4,4))
       )

### RNA Plot wiht nucleotides and dots ###

RNAPlot(dat)
RNAPlot(dat,nt=TRUE,add=TRUE,dp=0.75)
```

`rotateS` *Internal function to rotate a single point*

Description

Rotates a point a given angle around a given center point.

Usage

```
rotateS(x2, y2, x0, y0, ang)
```

Arguments

`x2`	x coordinate of the position being rotated
`y2`	y coordinate of the position being rotated
`x0`	x coordinate of the center of rotation
`y0`	y coordinate of the center of rotation
`ang`	rotatation angle in radians

Value

Returns a rotated point

Author(s)

JP Bida

Examples

```
### Rotate a point 90 degress ###
rotateS(0,1,0,0,pi/2)
```

`rotateV` *internal function to rotate a vector of points*

Description

Rotates a set of points around a center point a given number of radians

Usage

```
rotateV(x2, y2, x0, y0, ang)
```

Arguments

`x2`	Vector containing x coordinates being rotated
`y2`	Vector containing y coordinates being rotated
`x0`	x coordinate of center of rotation
`y0`	y coordinate of center of rotation
`ang`	Angle of rotation given in radians

Value

set of rotated points

Author(s)

JP Bida

See Also

`rotateS`

Examples

```
x=c(1,0,-1,0)
y=c(0,1,0,-1)
pts=rotateV(x,y,0,0,pi/4)
```

`stemCords` *internal function that generates coordinates for a stem*

Description

internal function that generates coordinates for an RNA secondary structure stem

Usage

```
stemCords(input, p1, p2, x1, y1, x2, y2, x3, y3)
```

Arguments

`input`	ct file as data frame
`p1`	index of nucleotide in first base pair of the stem
`p2`	index of nucleotide in first base pair of the stem
`x1`	x coordinate of p1
`y1`	y coordinate of p1
`x2`	x coordinate of p2
`y2`	y coordinate of p2
`x3`	direction vector x component
`y3`	direction vector y component

Value

set of points

Note

This is an internal function not recommend for use out side of the ct2coord function

Author(s)

JP Bida

See Also

`ct2coord`

Examples

```
### Internal Function ###
```

`transformFold` *Internal function to translate and rotate a secondary structure plot*

Description

Given a coordinate file, a point, and an angle in radians transformFold rotates the fold around the given point the given number of radians.

Usage

```
transformFold(dat, x0, y0, ang)
```

Arguments

`dat`	Coordinate file containing multiple RNA folds
`x0`	x coordinate of center of rotation
`y0`	y coordinate of center of rotation
`ang`	angle of rotation in radians

Value

dat frame containing the rotated coordinates

Author(s)

JP Bida

See Also

`alignCoord`

Examples

```
ct=makeCt("((((...(((((((.........)))))))...((((.........))))...))))",
          "AAAAAAAACCCCCCCCAAAGGGGGGGAUUACCCCUCCUUUAAAAGGGUUUUCCCCCCC")

c1=ct2coord(ct)

RNAPlot(c1)

c2=transformFold(c1,0,0,pi/2)
c3=transformFold(c2,0,0,pi/2)
c4=transformFold(c3,0,0,pi/2)

RNAPlot(c2,add=TRUE)
RNAPlot(c3,add=TRUE)
RNAPlot(c4,add=TRUE)
```

`translate` *internal function for translating points*

Description

internal function to translate points

Usage

```
translate(x1, y1, x2, y2)
```

Arguments

`x1`	x coordinates being translated
`y1`	y coordinates being translated
`x2`	dx for translation
`y2`	dy for translation

Value

set of points

Author(s)

JP Bida

Examples

```
## Internal Function ##
```

Index