

Package ‘EpitopePrediction’

March 3, 2017

Type Package

Title Predict Binding Between Peptides and MHC Molecules

Version 0.1-2

Date 2015-10-06

Maintainer Johannes Textor <johannes.textor@gmx.de>

Description Predicts binding of 9-to 12-mer peptides to MHC class I molecules using the stabilized matrix method. Can predict binding for several alleles from humans (HLA A/B), mice (H-2), chimpanzees (Patr A/B), and rhesus macaques (Mamu A/B).

VignetteBuilder knitr

LazyData true

License CC BY-NC-SA 4.0

Imports methods

Suggests knitr

RoxygenNote 5.0.1

Author Johannes Textor [aut, cre],
Bjoern Peters [aut]

R topics documented:

anchorPositions	2
binders	2
plotBindingMotif	3
smm	4
smmMatrix	5
supportedMHCs	5
Index	7

anchorPositions *Determine Anchor Positions*

Description

Uses the information in the SMM matrices to predict anchor positions for a given MHC molecule. Specifically, SMM matrix columns are sorted by their variance, and the k columns with lowest variance are defined as anchor positions.

Usage

```
anchorPositions(mhc = "HLA-A-02:01", l = 9, k = 2)
```

Arguments

mhc	the mhc molecules.
l	the peptide length.
k	how many anchor positions to return.

Examples

```
## Investigate anchor positions for the HLA molecules A02:01 and B27:05
anchorPositions( "HLA-A-02:01" )
anchorPositions( "HLA-A-02:01", k=3 )
anchorPositions( "HLA-B-27:05" )
```

binders *Find MHC Binders in Protein Sequence*

Description

Performs binding predictions for all l-mers in the given protein sequence, extracts the peptide whose IC50 values are below the given thresholds, and returns the results as a data frame.

Usage

```
binders(x, mhc = "HLA-A-02:01", l = 9, ic50.threshold = 500,
        quantile.threshold = NULL, include.peptide = TRUE, method = "smm")
```

Arguments

x	string, a protein sequence given in single-letter coding. Only the 20 common amino acids are supported.
mhc	string identifying the MHC molecule.
l	the peptide length.
ic50.threshold	peptides with a predicted IC50 value lower than this will be considered binders. A threshold of 500 nM is common. Use Inf to show predictions for all peptides of length l.

quantile.threshold	a number between 0 and 1. If this is not NULL, then the parameter ic50.threshold is ignored and the peptides for which the predicted IC50 falls within the given quantile will be returned. For instance, a value of .02 return the peptides whose binding strength is in the top 2% for the given protein.
include.peptide	logical, whether to include the actual peptide in the output data frame. This may not be desired in some circumstances, e.g. for very long proteins or for converting the result to a matrix.
method	string defining which prediction method to use. Currently the only implemented method is the stabilized matrix method (SMM), such that this setting is ignored. But further methods may be implemented in the future.

Value

a data frame containing the peptide (if include.peptide=TRUE), start position, end position, and predicted IC50 for every peptide below the threshold.

Examples

```
## This is the CORE protein from the Hepatitis C virus reference sequence available
## at https://hcv.lanl.gov/content/sequence/LOCATE/locate.html
hcv.core <- paste("MSTNPKPQRKTKRNTNRRPQDVKFPGGGQIVGGVYLLPRRGPRLGVRATRKTSERSQPRGR",
"QPIPKARRPEGRTWAQPGYPWPLYGNEGCGWAGWLLSPRGSRPSWGPTDPRRRSRNLGKVIDTLTCGFADLMGYIPLVGA",
"PLGGAARALAHGVRVLEDGVNYATGNLPGCSFSIFLLALLSCLTVPASA", sep="")
binders( hcv.core )
```

plotBindingMotif *Plot an MHC Binding Motif*

Description

Visualizes the set of peptides which are predicted to bind to a certain MHC molecule.

Usage

```
plotBindingMotif(mhc = "HLA-A-02:01", l = 9, motif.matrix = NULL,
width = 0.5, space = 1, col = rainbow(20), main = paste(mhc, " ", " ", l,
"-mers", sep = "" ), ...)
```

Arguments

mhc	name of the MHC molecule.
l	length of the peptide.
motif.matrix	a matrix. If this is not NULL, then the parameters mhc and l are ignored and the supplied matrix is used instead. This makes it possible to use this function for plotting sequence logos unrelated to MHC-peptide binding. The supplied matrix must have row names to indicate the letters out of which the sequence logo is made.
width	bar width for profile per position.
space	amount of space left before each bar, as a fraction of the bar width.

`col` vector of colors to use for the amino acids (given in alphabetical order in single-letter coding).

`main` title for the plot.

`...` further options to be passed on to [barplot](#)

Examples

```
## Compare binding motifs of HLA-A02 at all supported peptide lengths
par( mfrow=c(2,2) )
for( l in 8:11 ){
  plotBindingMotif( "HLA-A-02:01", l )
}
```

smm

Peptide-MHC Binding Prediction

Description

Predicts peptide-MHC binding using the stabilized matrix method (SMM) algorithm with a specifically constructed amino acid substitution matrix. See `citation(package='EpitopePrediction')` for the reference.

Usage

```
smm(x = c("SLYNTVATL", "SYFPEITHI"), mhc = "HLA-A-02:01",
    output.IC50 = TRUE)
```

Arguments

`x` vector of strings containing the peptides for which to predict MHC binding.

`mhc` string or vector of strings identifying the MHC molecules. See `supportedMHCs` for allowed values. If a vector is given, it must be of the same length as `x`.

`output.IC50` whether to output the IC50 value itself (default) or its base10-logarithm.

Examples

```
## Predict IC50 binding values for two famous peptides
smm( c("SLYNTVATL", "SYFPEITHI"), "HLA-A-02:01" )
```

smmMatrix

*Get Prediction Matrix***Description**

Loads the prediction matrix for a given MHC molecule at the given peptide length. Matrices are loaded from the package folder upon first load and are then cached in memory for further uses. The SMM method predicts binding by summing the matrix entries for each amino acid per position and then adding and MHC-dependant number. The resulting number is the log IC50 value (IC50=half maximal inhibitory concentration). Hence, low numbers mean strong binding and high numbers mean weak binding.

Usage

```
smmMatrix(mhc = "HLA-A-02:01", l = 9)
```

Arguments

mhc	name of the MHC molecule.
l	the peptide length.

Value

A list with components M (the matrix) and c (a number to be added to each prediction made with this matrix).

Examples

```
## load prediction matrix for HLA-A-02:01 9-mers
M <- smmMatrix( "HLA-A-02:01", 9 )$M
## How well does Leucine bind at each position?
M["L",]
## Which amino acid is most preferred at the C-terminal position?
names( which.min( M[,9] ) )

## Naming of the MHC molecule is permissive
for( mhc in c("HLA-A*0201", "HLA-A-0201", "HLA-A*0201", "HLA-A*02:01") ){
  names( which.min( smmMatrix( "HLA-A-02:01", 9 )$M[,9] ) )
}
```

supportedMHCs

*Supported MHC Molecules and Peptide Lengths***Description**

Supported MHC Molecules and Peptide Lengths

Usage

```
supportedMHCs(l = NULL)
```

Arguments

l integer giving the desired peptide length. If not given (default), then all possible combinations of peptide lengths and MHC molecules are returned.

Value

A data frame containing the combinations of supported MHCs and peptide lengths.

Examples

```
## Which MHC molecules are supported?  
levels( supportedMHCs()$mhc )  
## Which peptide lengths are supported for HLA-A02:01?  
with( supportedMHCs(), l[mhc=="HLA-A-02:01"] )
```

Index

anchorPositions, [2](#)

barplot, [4](#)

binders, [2](#)

plotBindingMotif, [3](#)

smm, [4](#)

smmMatrix, [5](#)

supportedMHCs, [5](#)