

Package: QSARdata (via r-universe)

September 19, 2024

Type Package

Title Quantitative Structure Activity Relationship (QSAR) Data Sets

Version 1.02

Date 2011-02-08

Author Max Kuhn

Maintainer Max Kuhn <mxkuhn@gmail.com>

Description Molecular descriptors and outcomes for several public domain data sets

Depends R (>= 2.10)

License GPL

LazyLoad yes

URL <http://qsardata.r-forge.r-project.org/>

Repository <https://r-forge.r-universe.dev>

RemoteUrl <https://github.com/r-forge/qsardata>

RemoteRef HEAD

RemoteSha 47aa3c1527c31d2da089238d563d37a2add95e94

Contents

| | |
|------------------------|----------|
| AquaticTox | 2 |
| bbb2 | 3 |
| MeltingPoint | 4 |
| Mutagen | 5 |
| Index | 6 |

AquaticTox

Fathead Minnow Acute Aquatic Toxicity

Description

These data were compiled and described by He and Jurs (2005). The data set consists of 322 compounds that were experimentally assessed for toxicity. The outcome is the negative log of activity (but is labeled as "activity"). The structures and outcomes were obtained from <http://www.qsarworld.com/index.php>.

The package contains none sets of molecular descriptors: atom pair distances, Daylight fingerprints (<http://www.daylight.com/dayhtml/doc/theory/theory.finger.html>), Dragon descriptors (http://www.taletе.mi.it/products/dragon_plus.htm), MOE2D, MOE2D fingerprints, MOE3D, PipelinePilot fingerprints (<http://accelrys.com/products/pipeline-pilot/>) and QuickProp descriptors (<http://www.schrodinger.com/products/14/17/>).

For fingerprints, the 500 most variable bits were selected whenever possible.

Usage

```
data(AquaticTox)
```

Format

The data consist of several data frames. The first column of the descriptor data frames is called "Molecule" representing the compounds.

AquaticTox_AtomPair Atom pair descriptors

AquaticTox_Daylight_FP Daylight fingerprints (<http://www.daylight.com/dayhtml/doc/theory/theory.finger.html>)

AquaticTox_Dragon Dragon descriptors (http://www.taletе.mi.it/products/dragon_plus.htm)

AquaticTox_Lcalc LCalc descriptors

AquaticTox_moe2D 2 dimensional MOE descriptors

AquaticTox_moe2D_FP 2 dimensional MOE fingerprints

AquaticTox_moe3D 3 dimensional MOE descriptors

AquaticTox_PipelinePilot_FP PipelinePilot fingerprints (<http://accelrys.com/products/pipeline-pilot/>)

AquaticTox_QuickProp QuickProp descriptors

AquaticTox_Class a factor with levels "Crosses" and "DoesNot"

AquaticTox_Outcome a data frame with columns for the molecule name and the outcome (for merging)

References

He and Jurs. Assessing the reliability of a QSAR model's predictions. *Journal of Molecular Graphics and Modelling* (2005) vol. 23 (6) pp. 503-523

Examples

```
data(AquaticTox)
head(AquaticTox_Outcome)
```

bbb2

Blood-Brain Barrier Data

Description

These data were compiled and described by Burns et al. (2004). The data set consists of 80 compounds that were designated as either crossing the blood-brain barrier or not crossing. The structures and outcomes were obtained from <http://www.qsarworld.com/index.php>.

The package contains none sets of molecular descriptors: atom pair distances, Daylight fingerprints (<http://www.daylight.com/dayhtml/doc/theory/theory.finger.html>), Dragon descriptors (http://www.taletе.mi.it/products/dragon_plus.htm), MOE2D, MOE2D fingerprints, MOE3D, PipelinePilot fingerprints (<http://accelrys.com/products/pipeline-pilot/>) and QuickProp descriptors.

For fingerprints, the 500 most variable bits were selected whenever possible.

There are compounds with missing data for some descriptors.

The "2" in the name is due to another data set in the **caret** package for blood-brain barrier data (with numeric outcomes). These are a completely different set of compounds and have no connection.

Usage

```
data(bbb2)
```

Format

The data consist of several data frames. The first column of the descriptor data frames is called "Molecule" representing the compounds.

bbb2_AtomPair Atom pair descriptors

bbb2_Daylight_FP Daylight fingerprints (<http://www.daylight.com/dayhtml/doc/theory/theory.finger.html>)

bbb2_Dragon Dragon descriptors (http://www.taletе.mi.it/products/dragon_plus.htm)

bbb2_Lcalc Lcalc descriptors

bbb2_moe2D 2 dimensional MOE descriptors

bbb2_moe2D_FP 2 dimensional MOE fingerprints

bbb2_moe3D 3 dimensional MOE descriptors

bbb2_PipelinePilot_FP PipelinePilot fingerprints (<http://accelrys.com/products/pipeline-pilot/>)

bbb2_QuickProp QuickProp descriptors

bbb2_Class a factor with levels "Crosses" and "DoesNot"

bbb2_Outcome a data frame with columns for the molecule name and the outcome (for merging)

References

Burns et al. A mathematical model for prediction of drug molecule diffusion across the blood-brain barrier. *The Canadian Journal of Neurological Sciences* (2004) vol. 31 (4) pp. 520-527

Examples

```
data(bbb2)
head(bbb2_Outcome)
```

MeltingPoint

Melting Point Data

Description

Karthikeyan et al (2005) presented data where they used chemical descriptors to model the melting point of compounds (i.e. transition from solid to liquid state). They assembled 4401 compounds: 4126 for model training and 275 compounds as a final validation set. They calculated 2D and 3D MOE chemical descriptors.

Usage

```
data(MeltingPoint)
```

Format

The descriptors are contained in a data frame called `MP_Descriptors` and the melting points are in a numeric vector `MP_Outcome`. The original data set indicators are in a factor vector called `MP_Data` with levels "Test" and "Train"

References

Karthikeyan et al. General melting point prediction based on a diverse compound data set and artificial neural networks. *Journal of chemical information and modeling* (2005) vol. 45 (3) pp. 581-90

Examples

```
data(MeltingPoint)
head(MP_Descriptors)
```

Mutagen

Mutagenicity Data

Description

Kazius et al (2005) investigated using chemical structure to predict mutagenicity (the increase of mutations due to the damage to genetic material). An Ames test was used to evaluate the mutagenicity potential of various chemicals. There were 4,337 compounds included in the data set with a mutagenicity rate of 55.3%. Using these compounds, the **DragonX** software (<http://www.taletе.mi.it/>) was used to generate a baseline set of 1,579 predictors, including constitutional, topological and connectivity descriptors, among others. These variables consist of basic numeric variables (such as molecular weight) and counts variables (e.g., number of halogen atoms).

Usage

```
data(Mutagen)
```

Format

The descriptors are contained in a data frame called `Mutagen_Dragon` and the outcomes are in a factor vector `Mutagen_Outcomes` with levels "mutagen" and "nonmutagen"

References

Kazius et al. Derivation and validation of toxicophores for mutagenicity prediction. *Journal of medicinal chemistry(Print)* (2005) vol. 48 (1) pp. 312-320

Examples

```
data(Mutagen)
head(Mutagen_Dragon)
```

Index

* datasets

- AquaticTox, [2](#)
- bbb2, [3](#)
- MeltingPoint, [4](#)
- Mutagen, [5](#)

- AquaticTox, [2](#)
- AquaticTox_Activity (AquaticTox), [2](#)
- AquaticTox_AtomPair (AquaticTox), [2](#)
- AquaticTox_Class (AquaticTox), [2](#)
- AquaticTox_Daylight_FP (AquaticTox), [2](#)
- AquaticTox_Dragon (AquaticTox), [2](#)
- AquaticTox_Lcalc (AquaticTox), [2](#)
- AquaticTox_moe2D (AquaticTox), [2](#)
- AquaticTox_moe2D_FP (AquaticTox), [2](#)
- AquaticTox_moe3D (AquaticTox), [2](#)
- AquaticTox_Outcome (AquaticTox), [2](#)
- AquaticTox_PipelinePilot_FP (AquaticTox), [2](#)
- AquaticTox_QuickProp (AquaticTox), [2](#)

- bbb2, [3](#)
- bbb2_AtomPair (bbb2), [3](#)
- bbb2_Class (bbb2), [3](#)
- bbb2_Daylight_FP (bbb2), [3](#)
- bbb2_Dragon (bbb2), [3](#)
- bbb2_Lcalc (bbb2), [3](#)
- bbb2_moe2D (bbb2), [3](#)
- bbb2_moe2D_FP (bbb2), [3](#)
- bbb2_moe3D (bbb2), [3](#)
- bbb2_Outcome (bbb2), [3](#)
- bbb2_PipelinePilot_FP (bbb2), [3](#)
- bbb2_QuickProp (bbb2), [3](#)

- MeltingPoint, [4](#)
- MP_Data (MeltingPoint), [4](#)
- MP_Descriptors (MeltingPoint), [4](#)
- MP_Outcome (MeltingPoint), [4](#)
- Mutagen, [5](#)
- Mutagen_Dragon (Mutagen), [5](#)
- Mutagen_Outcome (Mutagen), [5](#)