

Supplementary Material for:

The U.S. Dairy Forage Research Center (DFRC) Condensed Tannin NMR Database

Wayne E. Zeller* and Paul F. Schatz

US Dairy Forage Research Center, Madison, WI, 53706, USA

A User Guide for the DFRC Condensed Tannin Database

The database is constructed using FileMaker Pro 10.0v3 and updated to FileMaker Pro 14.0v5 before web launching. The entries consist of data for the monomers, dimers, trimers and tetramers of condensed tannins found by searching the Reaxys database. In many cases, the NMR information from any one reference is incomplete.

Appearance

The figure below shows how a typical flavan-3-ol/condensed tannin oligomer is presented in the database.

1: Reaxys Number 4345755

2: Formula C₃₀H₂₆O₁₄

3: DP 2

4: Natural Product

5: NMR

6: Natural Product References
Pharmazie, 2003, vol. 58, # 6 p. 416 - 419 (Cistus albidus L. (Cistaceae) collected at Massiv d' Esterel, France)
Phytochemistry (Elsevier), 1991, vol. 30, # 3 p. 981 - 985 (Cistus incanus subsp. tauricus)

7: Names Gallo catechin-(4α->6)-gallo catechin, gallo catechin-(4α,6)-gallo catechin

8: Structure

9: Backbone GC46GC

10: NMR Reference
Compound 5
Janecki, Aneta; Kolodziej, Herbert
Molecules, 2010, vol. 15, # 10 p. 7139 - 7152

11: NMR Solvent methanol

12: SMILES
O[C@H]1CC2=C(O[C@@H]1C1=CC(O)=C(O)C(O)=C1)C=C(O)C([C@H]1[C@H](O)[C@H](OC3=C1C(O)=CC(O)=C3)C1=CC(O)=C(O)C(O)=C1)=C2O

13: Comments 2 rotamers. d4-methanol

14: IUPAC Name

15: NMR References
Molecules, 2010, vol. 15, # 10 p. 7139 - 7152

16: NMR Data Table

| | ABC | | DEF | | GHI | | JKL | |
|----|-----------------|----------------|-----------------|----------------|-----------------|----------------|-----------------|----------------|
| | ¹³ C | ¹ H |
| 2 | 84.3 | 4.18 | 82.8 | 4.52 | | | | |
| 3 | 73.7 | 4.52 | 68.5 | 3.84 | | | | |
| 4 | 38.6 | 4.38 | 28.0 | 2.4-2.8 | | | | |
| 4a | 107.1 | | 100.5 | | | | | |
| 5 | 154.8-158.8 | | 154.8-158.8 | | | | | |
| 6 | 96.3 | 5.80 | 110.3 | | | | | |
| 7 | 154.8-158.8 | | 154.8-158.8 | | | | | |
| 8 | 96.9 | 5.73 | 97.5 | 5.92 | | | | |
| 8a | 154.8-158.8 | | 154.8-158.8 | | | | | |
| 1' | 131.7 | | 131.7 | | | | | |
| 2' | 108.4 | 6.04 | 108.8 | 6.35 | | | | |
| 3' | 146.3 | | 146.7 | | | | | |
| 4' | 133.9 | | 133.9 | | | | | |
| 5' | 146.3 | | 146.7 | | | | | |
| 6' | 108.4 | 6.04 | 108.8 | 6.35 | | | | |

Fields

Reaxys Number (1)

The field in the upper left corner of the screen is the number of this compound in the Reaxys database. The same compound may have several Reaxys number associated with it. Each unique Reaxys number is a separate entry in this database.

Formula (2)

The field to the right of the **Reaxys Number** is the chemical formula of the compound.

DP (3)

The next field to the right designates the degree of polymerization for the database entry – the number of monomeric tannin units that make up the compound.

Natural Product (4)

If this box is checked, some (or all) of the references have information regarding the natural occurrence of the compound.

NMR (5)

If this box is checked, some (or all) of the references have information regarding the NMR spectra of the compound.

Natural Product References (6)

This scrolling text field lists the references containing information regarding the origin(s) of the compound. The natural source is listed after the reference. This field has not been filled in for some of the compounds with a long list of references.

Name (7)

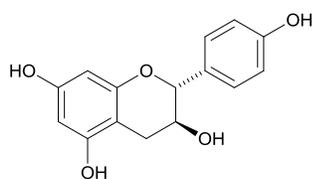
This scrolling text field lists the names of the compounds found in the references. In some cases, there are names listed that are inconsistent with other names in the field.

Structure (8)

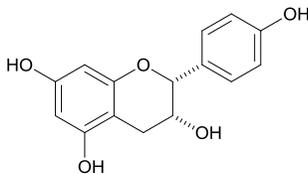
A line drawing of the molecule is displayed in the structure box. The usual convention for relative stereochemistry is used, i.e., dashed lines indicate bonds extending below the plane of the paper and wedges indicate bonds extending above the plane of the paper.

Backbone (9)

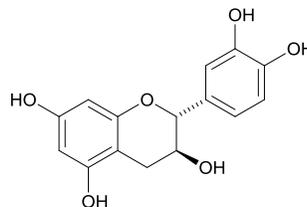
The backbone is a method for coding the structure to facilitate searching for partial or complete structures. The symbols for the common condensed tannin flavan-3-ol subunits are:



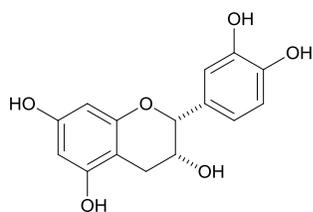
AF - Afzelechin



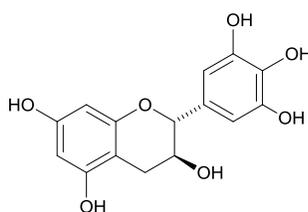
EF - Epiafzelechin



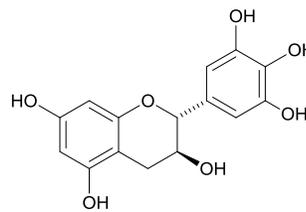
CA - Catechin



EC - Epicatechin

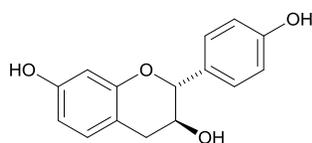


GC - Gallocatechin

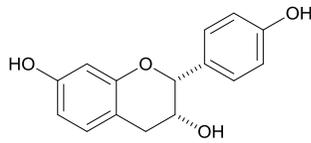


EG - Epigallocatechin

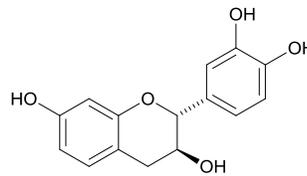
Symbols for some of the less common monomeric units are:



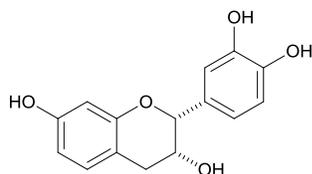
GB - Guibourtinidol



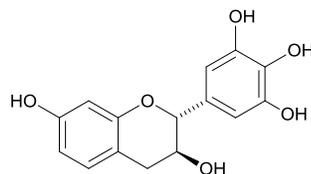
EB - Epiguibourtinidol



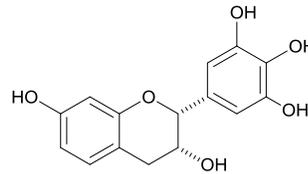
FS - Fisetinidol



EF - Epifisetinidol



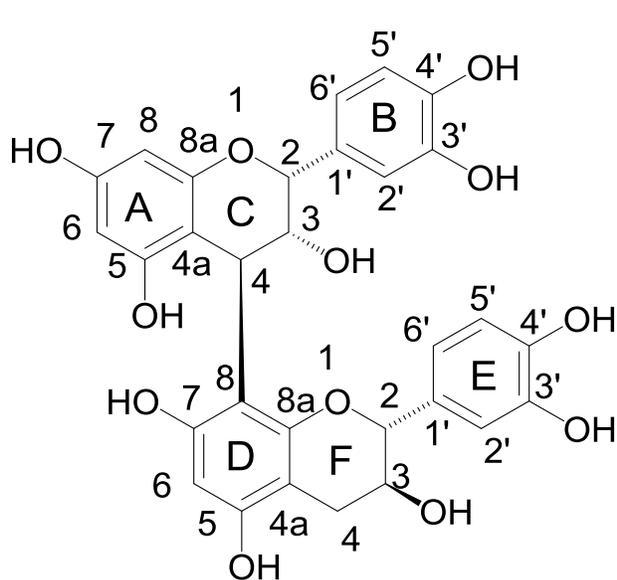
RB - Robinetinidol



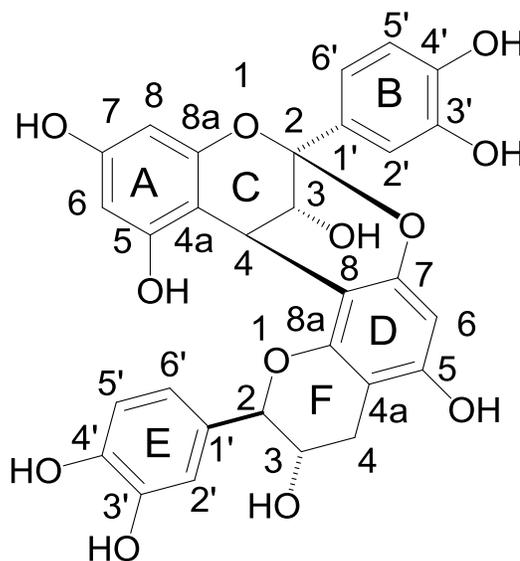
ER - Epirobinetinidol

The bonding between flavan-3-ol units is designated with the numbered atoms participating in the flavan-3-ol subunit of the oligomer. The link between the two tannin units below is designated 48. The 48 link and the 46 link are the most common types of single bond links between tannin units and are referred to as a B-type linkage.

Since the top tannin is an epicatechin unit and the bottom tannin is a catechin unit, the backbone code for this molecule is EC48CA.



EC48CA
B-Type Interflavan Linkage



EC482O7CA
A-Type Interflavan Linkage

Note that the relative stereochemistry of the hydroxyl group at position 3 of the heterocyclic ring of the top unit is *trans* to the link at position 4. This is usually the case.

For coding purposes, there is no differentiation between enantiomers of tannin units, i.e. both (-)-epicatechin and (+)-epicatechin are coded as EC.

The second major type of linkage is one in which two single bonds are formed between the flavan-3-ol subunits. This is demonstrated in the following structure and is referred to as A-type linkage. In this example, bond connectivity is coded as 482O7. The carbon atom at the 4 position of the top unit is singly bonded to the carbon atom at the 8 position of the lower unit and the carbon atom at 2 position of the upper unit is singly bonded to the oxygen atom attached to the carbon at the 7 position of the lower unit. The backbone code for this molecule is EC482O7CA. A-type linkages can also arise from the 46 linkage, i.e., 462O7.

NMR Reference (10)

The reference that provides the NMR data in the chemical shift matrix below the **Structure** box is listed. If there is an identifier for the molecule in the reference, it is included here.

NMR Solvent (11)

The solvent used to prepare the NMR sample is listed. All of the solvents are the deuterated analogs of the solvent, i.e., acetone = C₃D₆O, water = D₂O, methanol = CD₃OD, DMSO = C₂D₆SO, etc.

SMILES (12)

The SMILES (**S**implified **M**olecular **I**nput **L**ine **E**ntry **S**ystem) line notation of the molecule is listed. The description was generated with Marvin Sketch, a companion drawing application to Reaxys from Chem Axon. This field can be searched by placing any retrieved SMILES line notation from an outside source within quotation marks.

Comments (13)

This text box provides pertinent information about the molecule, such as the presence and number of rotamers detected, any corrections to chemical shifts, clarification of alphabetical ring designations, etc.

IUPAC Name (14)

To date, no IUPAC names have been added to the database.

NMR References (15)

This scrolling text box lists all of the references found that report NMR data for the molecule.

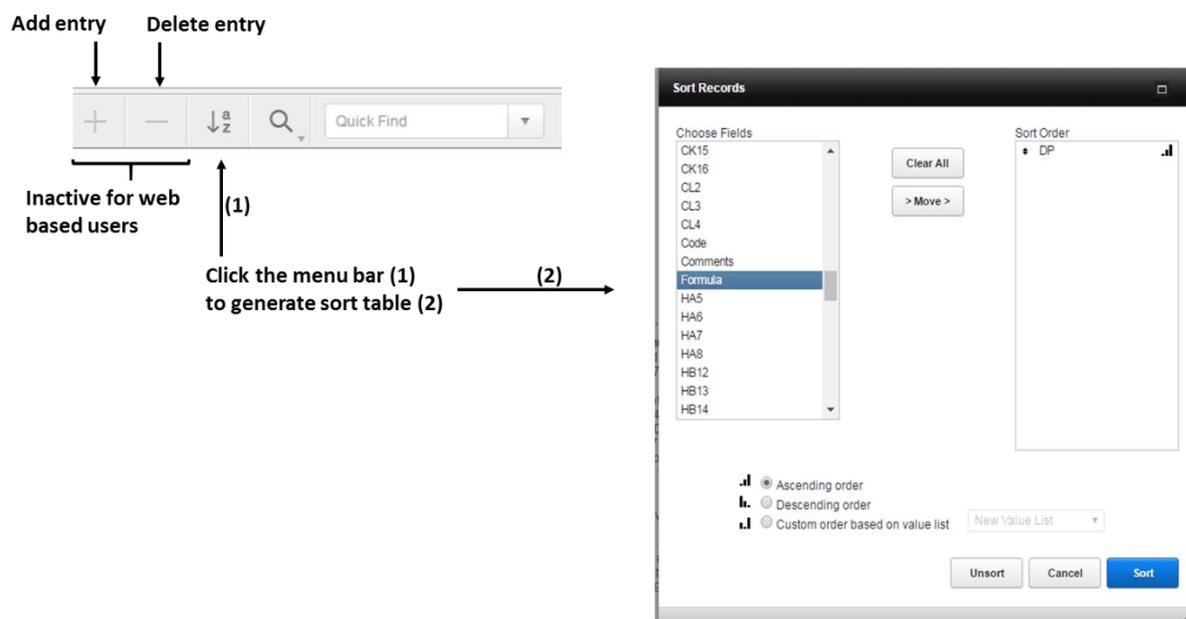
Chemical Shift Matrix (16)

The ¹H and ¹³C chemical shifts of the hydrogen and carbon atoms of the molecules are listed. The data is from the “best” or only data set found in the list of NMR references. The data is verbatim from the reference listed in the **NMR Reference field**. There are no modifications to account for differences due to different solvents or method of referencing. In many cases, the original reference reports ranges of chemical shifts.

The numbers along the left edge of the matrix refer to the number of the atom in the tannin unit (the previous structural figures show the typical numbering). The letters along the top of the matrix refer to the positioning of the tannin unit in a polymer. The ring labels follow previously used common designations to identify specific rings in CT oligomers. In this convention, sequential capital letters are used to identify first the phloroglucanol ring, followed by the phenyl ring projecting from C-2 and finally the dihydropyran ring of the monomeric unit. The letters ABC are reserved for the extender unit, followed in sequence (DEF) for the phloroglucinol, phenyl and dihydropyran rings, respectively, of the next extender unit. This capital letter labeling is continued until the terminal flavan-3-ol unit is reached. In a few cases, there is more than one possible terminal unit (i.e. entry 139, Reaxys number 20000364). In these cases, the capital letter labeling is clarified in the comment section.

Sorting

The records can be sorted according to one or more fields in a record. For example, to organize the records by degree of polymerization, click on the az icon in the toolbar. This will bring up a dialog box. The scrolling text box on the left of this dialog box displays all of the fields available for sorting. The right hand text box will be empty. If it is not empty, clicking on Clear All will remove any fields in the box. Click DP (degree of polymerization) in the left hand box and then click on Move. The right hand text box now displays DP. Click on DP in the right hand text box to highlight the entry, choose the order for sorting (usually ascending), and click on Sort.



The records may also be sorted according to multiple fields. The fields to be used for sorting are moved to the right hand text box in the same fashion as above. The order by which the fields are sorted is set by position of the fields in the text box. Click and drag a field to the requisite position. The top field in the text box is sorted first; the bottom field is sorted last.

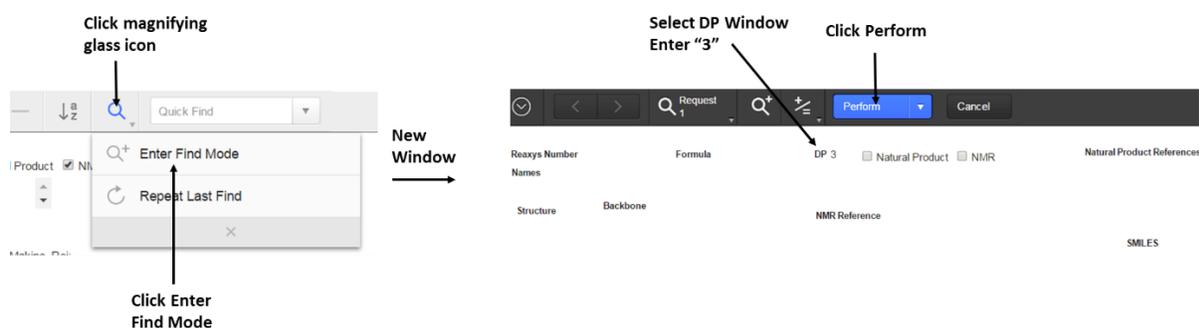
Searching

The search function of the database will generate a subset defined by criteria designated in one field or multiple fields. The subset can then be browsed, sorted, or searched to generate a subset of the original subset. In the toolbar above the record, to the right of the record number, is a pie chart providing a visual indication of the size of the subset relative to the full data set. The exact ratio of the number of records in the subset to the number of records in the full data set is

displayed to the right of the pie chart. To regenerate the full data set, click on the **Show All** icon in the toolbar.

Single Field Search

Enter the search mode by clicking on the magnifying glass icon in the toolbar above the record. A drop down menu appears and select **Enter Find Mode**. The data in the fields of the displayed record are replaced with magnifying glass icons. As an example, to generate a subset of trimers of tannin units, click the **DP** field and enter the number **3**. The search is initiated by clicking on the **Perform** icon that has appeared in the toolbar. Alternatively, pressing the *return* key will also initiate the search. The generated subset contains all of the trimeric tannin molecules in the database. This subset can be refined further by doing a search of it. The subset will contain entries that have natural product references and NMR references. To find the records with NMR references, click on the **Find** icon, click in the **NMR** field, and press the *return* key. The new subset will contain all of the trimeric tannins with NMR references.



Multiple Field Search

Two or more fields can be searched to generate a subset that matches criteria from all the fields. For example, to generate the subset from the previous example, click on the **Find** icon, enter **2** in the **DP** field, check the **NMR** field, and press the *return* key.

Backbone Search

There are multiple methods for searching the Backbone field. The use of special characters in the search field defines the subset generated by the search. The commonly used special characters are:

* - one or more unknown characters

- digit

== - exact match (two equal signs)

The table below provides examples of how the special characters affect the results.

| Characters Entered in Backbone Field | Records in Subset |
|--------------------------------------|---|
| EC | Epicatechin is the first (top) unit of the oligomer (includes monomer). |
| EC* | Epicatechin is the first (top) unit of the oligomer (excludes monomer). |
| *EC | Epicatechin is the last (bottom) unit of the oligomer (includes monomer). |
| *EC* | Epicatechin is an internal unit. |
| ==EC | Epicatechin monomer. |
| ==EC48CA | Epicatechin/catechin dimer (48 linkage). |
| *EC#6CA* | Internal epicatechin/catechin units with 46 linkage. |
| EC*CA | Epicatechin is the first unit and catechin is the last unit of the polymer. |

Chemical Shift Search

The chemical matrix can be searched with one or multiple fields. Entering a single value into a field generates a subset that only contains the number specified. For example, if 31.7 is entered in the DEF/4 field, only the records that contain 31.7 in that field are in the subset. Specifying a range of values is usually the preferred method for searching. To search a range, enter the first and last values of the range separated by two periods. For example, if 31.5 .. 32.5 is entered in the DEF/4 field, records that contain any value between 31.5 and 32.5 are included in the subset.