

# Quality Database for Pure Component Properties Estimation Methods

*Artist - Program Package for the Estimation of Pure Component  
Properties from Molecular Structures*

DDBST Software and Separation Technology GmbH

Marie-Curie-Straße 10

D-26129 Oldenburg

Tel. +49 (0) 441 361819 0

Fax: +49 (0) 441 361819 10

Mail: [support@ddbst.de](mailto:support@ddbst.de)

## Contents

Introduction.....	1
The Dortmund Data Bank (DDB) – Pure Component Properties.....	2
Artist – Estimation of Pure Component Properties from Molecular Structures.....	3
The Quality Data Bank.....	4
Classifying Components.....	4
Obtaining the Deviations.....	5
Using the Quality Data Bank in Artist.....	5
Available Quality Data.....	7

## Introduction

A large number of estimation methods for pure component properties have been published in literature. Many of these have been implemented in the program package Artist together with new methods developed using the DDB and Artist. This leads to the impression, that many old methods are now obsolete and new developments should be used instead.

The most important features of a method are its reliability (security against large errors), range of applicability, low mean error and ease of use.

Reliability and mean error are difficult to judge from the original publication. In many publications the number of example calculations is way too small. Often too few data are used to compare with – in many cases even the database used for the model development is too small to provide reliability even if the few regarded components are estimated accurately. Reliability might be seen as estimation quality for components not used in the model development.

With the availability of a large data bank with experimental pure component data (DDB-Pure) and the program package Artist DDBST GmbH has tools to overcome

these problems. It is now possible to test many published methods against the largest pure component properties database available.

The quality database described here uses a quite simple definition of quality: The lower the mean deviation is the higher is the quality. The quality database contains mean errors for the different models for individual component classes.

## The Dortmund Data Bank (DDB) – Pure Component Properties

The DDB is a large database for thermodynamic, transport, and caloric data. Besides the pure component properties it also contains many data for mixtures.

These data banks contain almost only experimental data from original sources plus few derived data. All data were evaluated manually and automatically.

The current status of the pure component part of the database is shown here:

<i>Property</i>	<i>Compounds</i>	<i>References</i>	<i>Sets</i>	<i>Points</i>
Viscosity	2211	2424	15285	96570
Vapor Pressure	6144	6303	23354	162393
Critical Data	957	885	3279	3285
Triple Point	234	180	294	294
Kinematic Viscosity	611	234	1413	6379
Density	7303	5975	39407	288476
Melting Point	4832	2670	11399	12692
Molar Heat Capacity (cP)	2386	1831	9773	172556
Virial Coefficients	260	326	828	4575
Heat of Vaporization	2408	1042	4883	11056
Heat of Fusion	1423	837	2060	2105
Thermal Conductivity	771	909	8801	82022
Surface Tension	2054	635	4589	20220
Entropy	1460	899	2580	7975
Std. Heat of Combustion	1268	393	1471	1480
Std. Heat of Formation	2775	1191	3975	4009
Enthalpy (H-H0)	273	172	461	7126
Enthalpy (H-H298)	20	17	27	458
Gibbs Energy of Form./T	58	25	62	927
Gibbs Energy of Form.	33	25	36	104
G-function	1208	563	1896	28327
Enthalpy (H-H298)/T	3	2	3	76
Enthalpy (H-H0)/T	96	47	113	1892
Transition Temperature	383	331	755	855

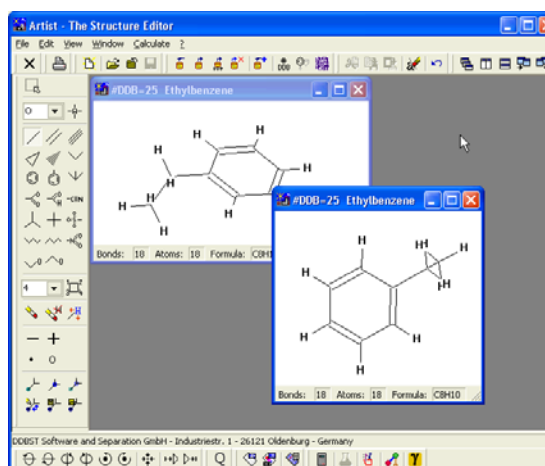
<i>Property</i>	<i>Compounds</i>	<i>References</i>	<i>Sets</i>	<i>Points</i>
Heat of Transition	360	305	628	643
Molar Heat Capacity (cV)	84	79	931	10382
Mass Heat Capacity (cV)	7	4	15	474
Ideal Gas Heat Capacity	1285	708	2251	30827
Dielectric Constant	334	114	589	3252
Speed of Sound	474	274	2392	20286
Diffusion Coefficient	8	3	22	129
Flash Point	1	1	1	1
Molar Saturation Heat Capacity	30	10	66	1051
Heat of Sublimation	14	2	14	14
Enthalpy of Sublimation	69	19	116	222
Entropy of Transition	2	1	2	4
Entropy of Formation	4	2	4	4
Enthalpy	47	18	77	3776
Entropy (S-S0)	2	2	2	85
Entropy (S-S298)	1	1	1	8
Thermal Expansion Coefficient	26	3	50	334
Compressibility (isothermal)	2	1	14	154
<b>Total</b>	<b>14834</b>	<b>18898</b>	<b>143919</b>	<b>987498</b>

This database is now used for checking the quality of the methods implemented in the Artist program package.

## Artist – Estimation of Pure Component Properties from Molecular Structures

This program package contains a structure editor, a database for molecular structures, and a calculation module.

The structure database currently contains more than 18300 molecular structures and approx. 75 mainly group contribution models are implemented in the calculation module.



## The Quality Data Bank

The quality database does not contain information for single components. Instead of a component-based approach the database uses chemical classes based on functional groups like alcohols, ketones, ethers, aromatic compounds, ....

Components are assigned to chemical classes and mean errors for the component classes are stored. The classification scheme was developed during our work on new group contribution methods. Very often simple component classes contain subclasses that need to be distinguished by an estimation method. In case of the C=O

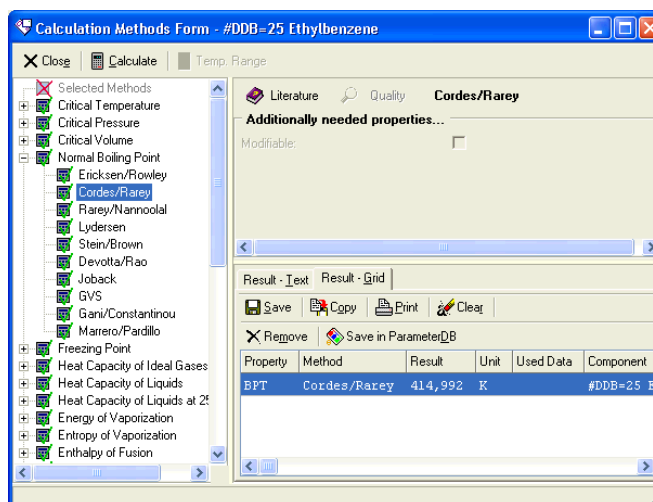
group found in carboxylic acids, esters, aldehydes and ketones, most physical properties are significantly influenced by a conjugated C=C-bond. In this case (C=C-C=O) the strongly electronegative oxygen is able to polarize the whole conjugated pi-electron system, which leads to a significantly larger dipole moment. The classification scheme therefore needs to distinguish between the different subgroups. While a method might yield good results for saturated carboxylic acids, it may lead to large deviations when applied to carboxylic acids with a conjugated double-bond.

Special attention was paid to components with more than one hydrogen-bonding group. These groups (like for example the alcohol-group) are not additive in a group contribution method and nearly all methods fail when applied to e.g. diols.

### Classifying Components

The classification of components is based on the work by Rarey/Nannoolal<sup>1</sup>. The model presented in that article is a group contribution method allowing to estimate the normal boiling point of a component from molecular structure only. The model specific groups are obtained from an automatic fragmentation algorithm implemented in the Artist package. The normal boiling is calculated as a function of the sum of group specific constants by a rather simple equation.

The group list used in that model is used for the classification of components. The component classes are not directly related to the groups since the fragmentation is in some cases too detailed, in other cases not specific enough and also doesn't represent commonly known classes very well.



<sup>1</sup> Nannoolal Y., Rarey J., Ramjugernath D., Cordes W., "Estimation of Pure Component Properties. Part 1. Estimation of the Normal Boiling Point of Non-Electrolyte Organic Compounds via Group Contributions and Group Interactions", Fluid Phase Equilib., 226, p45-63, 2004

For that reason a list of rules has been introduced. A rule assigns components with special groups (from the normal boiling model) to chemical classes like ketones and hydrocarbons. Additionally it allows or automatically discard unwanted combinations of groups.

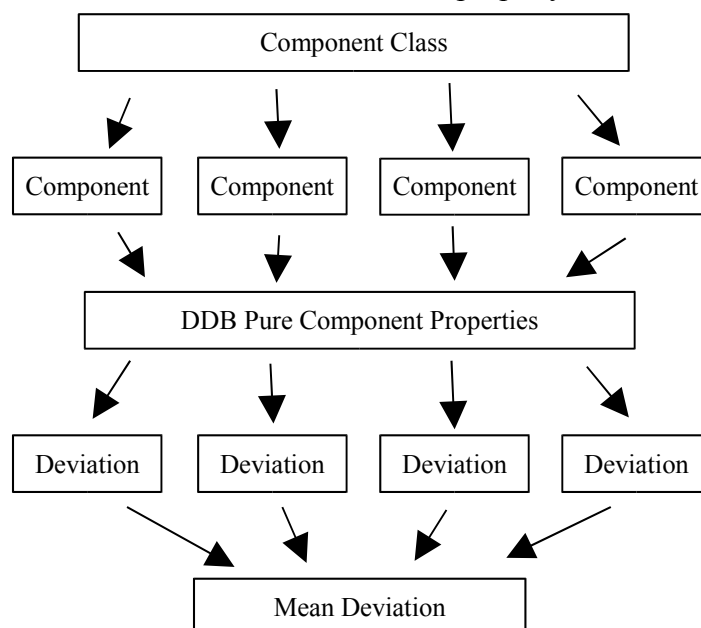
Examples:

<i>Class</i>	<i>Description</i>
Aromatic hydrocarbons	Only aromatic C and CH groups are allowed. No other group may be present.
Unsaturated hydrocarbons	Only non aromatic C, CH, CH <sub>2</sub> , CH <sub>3</sub> groups are allowed. At least a single double bond must be given
Aliphatic esters	Only non aromatic C, CH, CH <sub>2</sub> , CH <sub>3</sub> and the C=O-O-C (ester) groups are allowed.

In total the list of rules comprises rules for 187 component classes.

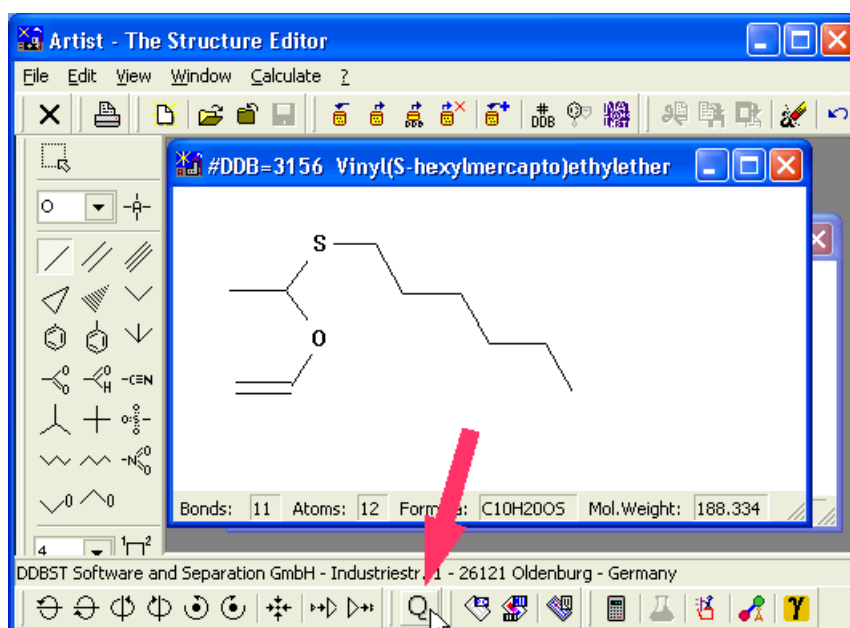
### ***Obtaining the Deviations***

The deviations are obtained by recalculating the entire DDB pure component properties (PCP) database component by component separated by component classes and storing the obtained mean deviations for a property and model.



### **Using the Quality Data Bank in Artist**

The quality data bank is an integrated part of Artist and can be used by selecting buttons in the structure editor. The dialog for displaying the quality information shows the component name and the DDB number of the component if available.



BPT Models Deviation

Close Copy Print

Component Vinyl(S-hexylmercapto)ethylether

DDB Number 3156

EtherO - Teth			
	AAD	Component Count	Quality
Rarey/Nannoolal	7.26 K	3	Good
Cordes/Rarey	15 K	3	Unreliable
Devotta/Rao	16.23 K	3	Unreliable
Joback/Reid	17.47 K	3	Unreliable
Stein/Brown	19.52 K	3	Unreliable
Ether			
	AAD	Component Count	Quality
Rarey/Nannoolal	7.4 K	190	Good
Cordes/Rarey	8.46 K	190	Good
Marrero/Pardillo	9.81 K	132	Good
Stein/Brown	11.89 K	163	Unreliable
Devotta/Rao	15.09 K	162	Unreliable
Joback/Reid	15.49 K	162	Unreliable
Gani/Constantinou	15.98 K	126	Unreliable
EWOR	17.36 K	148	Unreliable

Teth, Nitro, AtS, CN, AO, AN5, AN6 AAD Component Count Quality

The yellow lines identify the component class. The tables for the single classes contain the average absolute deviation (AAD), the component count originally used to obtain the AAD and a statement of quality. This example shows that a component can be part of multiple component classes. For the user it is always a good idea to go through the list and find the worst average absolute error and judge the quality of the model by this value.

## Available Quality Data

Data banks are available for

- normal boiling points
- critical temperatures
- critical pressures
- critical volumes

The list of data banks will be extended in the future.