

Quantitative Estimate of Drug-likeness in StarDrop

In a recent article, Bickerton *et al.* (2012) "The Chemical Beauty of Drugs" *Nature Chemistry* **4**, 90–98 (the QED paper), the authors proposed a measure of 'drug-likeness', the Quantitative Estimate of Drug-likeness (QED), that relates the similarity of a compound's properties to those of oral drugs based on eight commonly used molecular properties:

- Molecular weight (MW)
- Lipophilicity (logP)
- Number of hydrogen bond donors (HBD)
- Number of hydrogen bond acceptors (HBA)
- Polar surface area (PSA)
- Number of rotatable bonds (ROTB)
- Number of aromatic rings (AROM)
- Count of alerts for undesirable substructures (ALERTS)

The QED is based on a method for multi-parameter optimisation known as 'desirability functions'. A desirability function relates the value of a compound characteristic to the 'desirability' of that outcome. The desirability is a number between zero and one, where a value of one indicates that the outcome is ideal and a value of zero indicates that the outcome is completely unacceptable. Desirability functions are equivalent to the scoring functions used in StarDrop's Probabilistic Scoring approach, making this approach very amenable to implementation in StarDrop.

To derive the QED metric, desirability functions were fitted to the distributions of the eight properties listed above for 771 marketed oral drugs. Using this method, a higher desirability score is assigned to a compound for a given property if the compound's property is commonly observed amongst marketed oral drugs. An example is shown in Figure 1 for molecular weight. The desirabilities of all of the individual properties are combined into a single score, the QED, by taking their geometric mean.

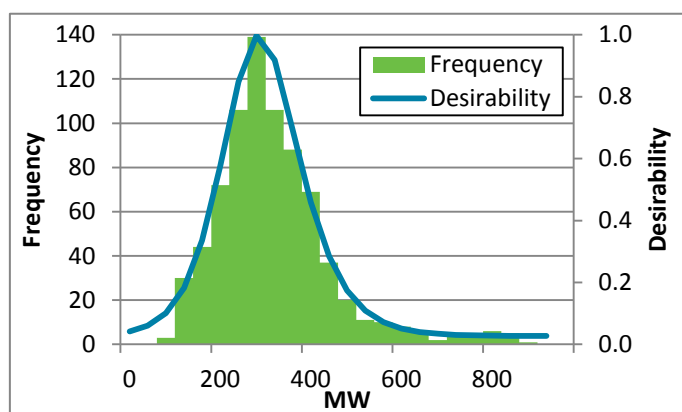


Figure 1 This graph shows the distribution of MW for a set of 771 orally absorbed small molecule drugs and a desirability function (blue), as used in QED, fitted to this distribution. The most desirable property values correspond to those most frequently observed in the set of drugs.

In the QED paper, the authors showed that the QED performed well in identifying a set of 771 marketed oral drugs, taken from DrugBank [11], from a set of 10,250 small-molecule ligands from the Protein Data Bank (PDB) ligand dictionary [12] (note that this was a different set of 771 oral drugs from that used to fit the desirability functions, although there was some overlap). Furthermore, the authors showed that the QED values agreed with medicinal chemists' subjective views of the attractiveness of compounds as hits on which to undertake further chemistry.

The QED can be conveniently represented as a scoring profile in StarDrop and in this directory we provide example profiles, along with calculators for the relevant descriptors. A detailed description of the contents of the directory and how to use them is given below.

Scoring Profiles

We have provided two scoring profiles, described in the following sections. We also provide instructions below on how to use these within StarDrop.

QED PP Properties

The scoring profile **QED PP Properties** calculates the QED value based on compound properties calculated in Pipeline Pilot, as used in the QED paper. The authors of the QED paper provide an example protocol to calculate the eight properties in the [supplementary information](#) to the paper.

The scoring function for each property was defined as a series of linear splines that approximates the form of the corresponding desirability function, as illustrated in Figure 2 for MW.

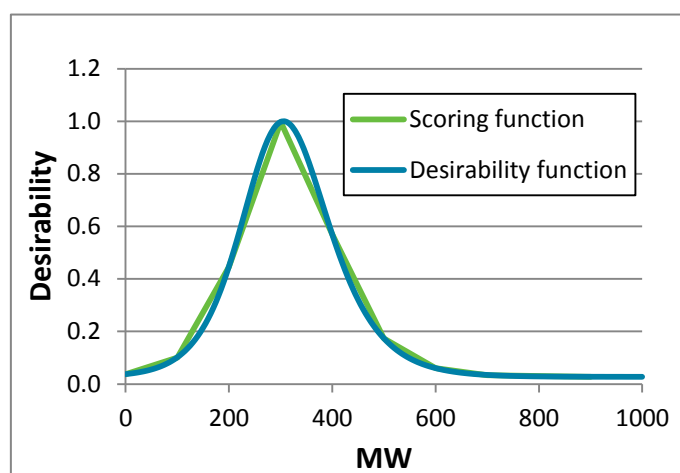


Figure 2 Example of fit of the MW scoring function to the QED desirability function using linear splines.

The unweighted QED value proposed by Bickerton *et al.* is defined as the geometric mean of the individual property desirabilities, while StarDrop's scoring algorithm is the product of the desirability values (in the absence of uncertainty). Therefore, the QED can be calculated from the score by taking the 8th root. This can be easily achieved using the function editor in StarDrop, as described below. The resulting QED values agree well with those provided in the supplementary information to the QED paper for 771 oral drugs, as shown in Figure 3. These data are provided in the file **QED PP Properties.add**.

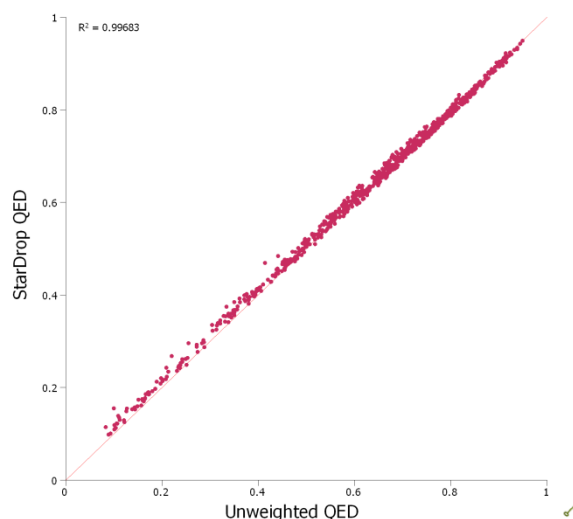


Figure 3 Graph showing the correlation between unweighted QED, as calculated in the QED paper, and QED calculated within StarDrop using property values calculated in Pipeline Pilot and the 'QED PP Properties' scoring profile. The coefficient of determination is 0.997, indicating a very good agreement.

QED StarDrop Properties

The profile **QED StarDrop Properties** uses descriptors calculated directly within StarDrop to calculate the QED. Some of the properties calculated in Pipeline Pilot do not correspond exactly with those calculated in StarDrop (or elsewhere) and therefore, where necessary, new models have been developed to calculate these properties. Unfortunately, Accelrys do not publish exact definitions of the descriptors generated by Pipeline Pilot, so these models have been generated by trial and error to produce as good an agreement as possible. Table 1 shows the models used in the QED StarDrop Properties profile, along with their agreement with the properties calculated in the QED paper using Pipeline Pilot.

Table 1 Models used within StarDrop to calculate the QED properties, showing the agreement with the values published in the supplementary information to the QED paper. The R^2 is the coefficient of determination of the property values reported with the QED paper and those calculated in StarDrop for a data set of 766 oral drugs. For integer counts, the percentage of compounds with identical values is also shown.

Property	Model	Description	Agreement with QED Property
MW	MW	Standard StarDrop MW calculator	$R^2=1.000$
logP	logP	Standard StarDrop logP model	$R^2=0.869$
HBD	QED_HBD	SMARTS-based model	$R^2=0.987$ (98% identical)
HBA	QED_HBA	SMARTS-based model	$R^2=0.977$ (88% identical)
PSA	QED_PSA	Topological surface area including N, O, S and P according to Ertl <i>et al.</i> [4]	$R^2=0.999$
ROTB	QED_ROT B	SMARTS-based model	$R^2=0.994$ (96% identical)
AROM	QED_AROM	Count of aromatic rings. Variations are due to differences in aromaticity perception.	$R^2=0.893$ (91% identical)
ALERT	QED_ALERTS	Count of alerts based on SMARTS defined in [5]. Variations are mainly due to differences in aromaticity perception.	$R^2=0.842$ (94% identical)

The QED-related models are provided in this directory and can be loaded into StarDrop as described below.

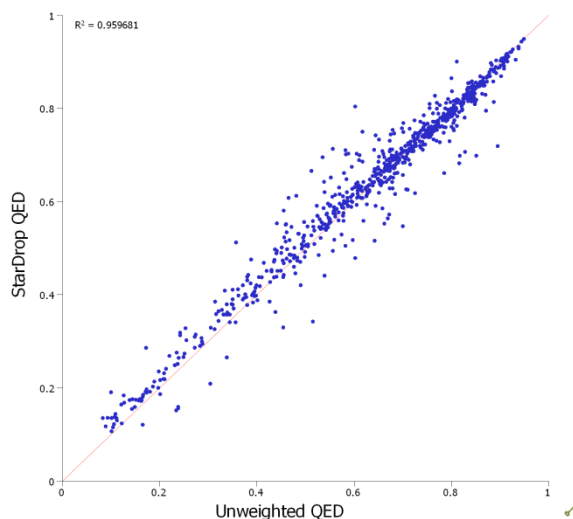



Figure 4 Graph showing the correlation between unweighted QED, as calculated in the QED paper, and QED calculated within StarDrop using property values calculated in StarDrop and the 'QED StarDrop Properties' scoring profile. The coefficient of determination is 0.960, indicating a very good agreement.

As noted above, the QED value can be calculated by taking the 8th root of the score. The results of applying the QED StarDrop Properties scoring profile to 766 compounds (a subset of the 771 compounds published in the QED paper for which reliable structures could be obtained) is shown in Figure 4. This also shows a good agreement with the published QED values for these compounds, with a coefficient of determination of 0.960. The variations between the QED values calculated using StarDrop properties and those calculated using Pipeline Pilot properties is due to two factors: the small differences in the calculated descriptors, as described in Table 1 and the uncertainty in the calculated logP, which is explicitly taken into account by StarDrop's probabilistic scoring. These data are provided in the file **QED StarDrop Properties.add**.

Calculating QED in StarDrop

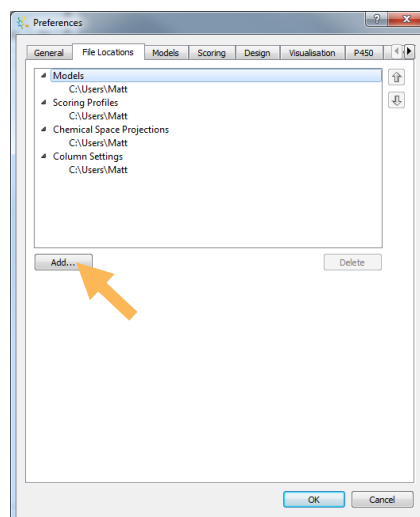
To start, save the contents of this directory in a convenient location (in the examples below we will use a directory called **StarDrop QED**).

Loading the StarDrop Property Calculators

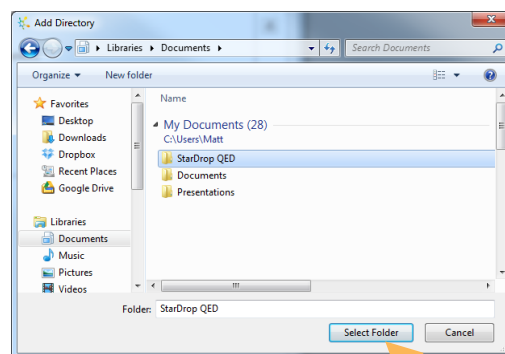
If you wish to calculate the compounds properties on which the QED is based in StarDrop the additional models should be loaded into StarDrop. These can be loaded individually using the  button on the **Models** tab.

However, it may be more convenient to add all of the models in the directory to those that are loaded automatically whenever StarDrop is started. To do this, select the **File->Preferences** menu option and change to the **File Locations** tab in the **Preferences** dialogue, as shown to the right.

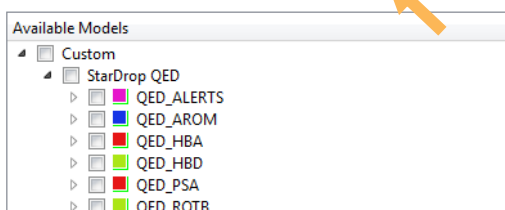
Select **Models** and click **Add**.



Navigate to the location of the **StarDrop QED** folder, select this and click **Select Folder** as shown to the right.



Finally, click **OK** in the Preferences dialogue and the models in the **StarDrop QED** directory will be added to the list of models under **Custom** in the **Models** tab.



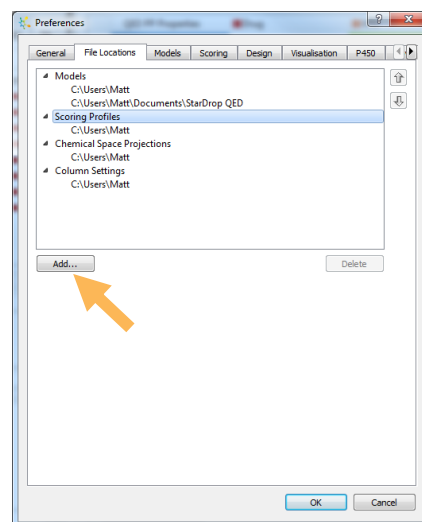
Loading Pre-calculated Property Values

If you have pre-calculated the property values, for example in Pipeline Pilot, you can load these from a comma-separated value (CSV), SD or text delimited file using the **File->Open** menu option. The property names in the file header should correspond to the property abbreviations listed above and used in the **QED PP Properties** scoring profile. An example of a suitable CSV file is provided in file **QED PP Properties Example.csv**.

Using the Scoring Profiles

The scoring profiles can be loaded using the  button on the **Scoring** tab. When loaded, the scoring profile will be listed under **Saved profiles** on the **Scoring Tab** and, when selected, the profile will appear in the profile view, as shown on the next page.

Alternatively, it may be more convenient to automatically load these profiles whenever StarDrop starts. This can be configured in a similar manner to the StarDrop property calculators. Select the **File->Preferences** menu option and change to the **File Locations** tab in the **Preferences** dialogue, as shown to the right.



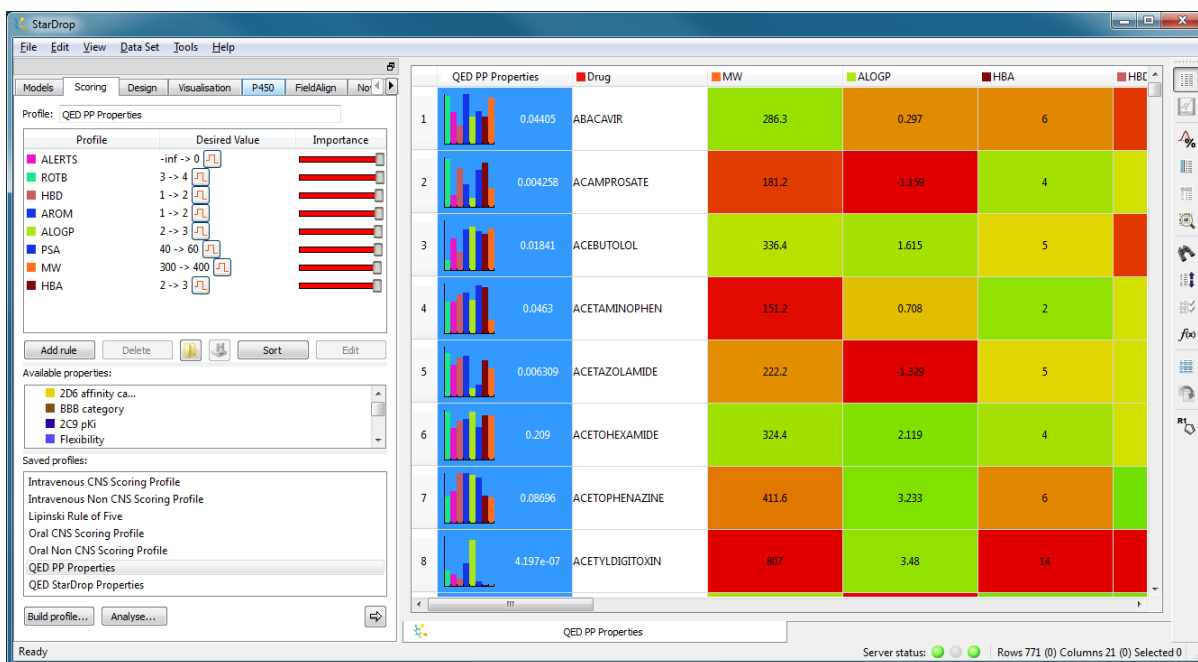
Select **Scoring Profiles** and click **Add**.

Navigate to the location of the **StarDrop QED** folder, select this folder and click **Select Folder**

Finally, click **OK** in the Preferences dialogue and the scoring profiles in the **StarDrop QED** directory will be added to the list of **Saved profiles** in the **Scoring** tab, as shown on the next page.

To run a profile, select the profile from the Saved profiles list and click the  button on the **Scoring** tab.

If you are using the QED StarDrop Properties profile, the properties will automatically be calculated (if the property calculators have been loaded as described above). You will get a warning that some of the columns contain values with zero uncertainty and this is correct; properties such as MW and counts of substructures don't have any statistical error and you can simply click **OK**.



The scores for the compounds in the data set will be calculated and displayed, as shown above. The histograms in the score column indicate the contribution of each property to the overall score and this information is also shown as a heat map when the scoring column is selected.

Calculating the QED Values

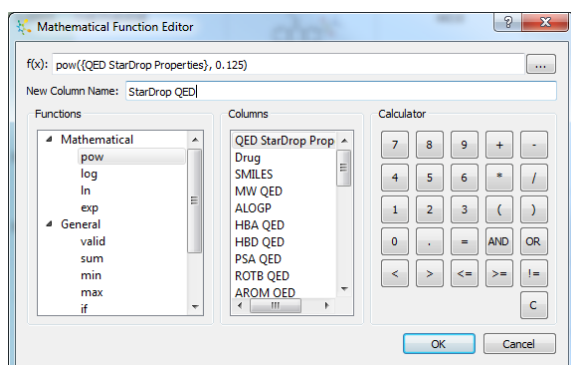
The QED is defined as the geometric mean of the desirabilities of the individual properties, while the StarDrop probabilistic scores correspond to the product of the individual desirabilities (in the absence of uncertainty). Therefore, the QED value can be easily calculated from the scores by taking the 8th root, using StarDrop's mathematical function editor.

To do this, after calculating the scores, click the $f(x)$ tool on the toolbar to open the mathematical function editor and enter one of the following formulae in the $f(x)$ field:

For scores generated with the QED PP Properties profile: $\text{pow}(\{\text{QED PP Properties}\}, 0.125)$

For scores generated with the QED StarDrop Properties profile: $\text{pow}(\{\text{QED StarDrop Properties}\}, 0.125)$

Hint: you can copy and paste this from above or simply point and click to enter the formula in the editor. An example is shown below:



Click **OK** and a column with the name entered in the **New Column Name** field will be created in your data set containing the calculated QED values.

Directory Contents

This directory contains the following files:

- **QED for StarDrop.pdf**: This document.
- Scoring Profiles
 - **QED PP Properties.apd**: A scoring profile for calculation of QED based on compound properties generated in Pipeline Pilot.
 - **QED StarDrop Properties.apd**: A scoring profile for calculation of QED based on compound properties calculated in StarDrop
- Data Sets
 - **QED PP Properties.add**: An example StarDrop file containing results of calculations of QED based on compound properties generated in Pipeline Pilot for 771 oral drugs. The results published in the QED paper are also provided for comparison.
 - **QED StarDrop Properties.add**: An example StarDrop file containing results of calculations of QED based on compound properties generated in StarDrop for 766 oral drugs. The results published in the QED paper are also provided for comparison.
 - **QED PP Properties Example.csv**: An example CSV file containing structures and compound properties calculated in Pipeline Pilot for 164 candidate compounds intended for GPCR targets.
- StarDrop Models
 - **QED_HBA.aim**: A StarDrop model that calculates the number of hydrogen bond acceptors for use in calculating the QED.
 - **QED_HBD.aim**: A StarDrop model that calculates the number of hydrogen bond donors for use in calculating the QED.
 - **QED_PSA.aim**: A StarDrop model that calculates the polar surface area for use in calculating the QED.
 - **QED_ROT.aim**: A StarDrop model that calculates the number of rotatable bonds for use in calculating the QED.
 - **QED_AROM.aim**: A StarDrop model that calculates the number of aromatic rings for use in calculating the QED.
 - **QED_ALERTS.aim**: A StarDrop model that calculates the number of structural alerts for use in calculating the QED.