

## FAQ

### StarDrop: Chemical Space

Connecting the chemical variations in a compound set with the properties or scores of those compounds is essential to understanding multi-dimensional structure-activity relationships. The distribution of properties and scores across the chemical space of your compounds can be visualized in StarDrop. The result is a two-dimensional scatter plot that relates the similarity of two compounds to the distance between the points representing those compounds. This enables the diversity of a group of compounds to be qualified relative to the overall diversity of the set from which they were selected.

The Chemical Space visualization within StarDrop allows the user to:

- Create a chemical space projection based on molecular structure ('chemical structure' space plot), the properties in the data set ('property' space plot) or a combination of the two ('mixed' space plot).
- Plot multiple data sets in one chemical space projection - a useful tool when comparing different chemical series within a project.
- View the structure of each compound in the plot by pointing at the corresponding point.
- Select compounds from the chemical space plots for more detailed inspection or progression to further analysis.
- Visualize trends by coloring the dots in the chemical space plot on a user-defined scale according to any property or probabilistic score.

## Frequently Asked Questions

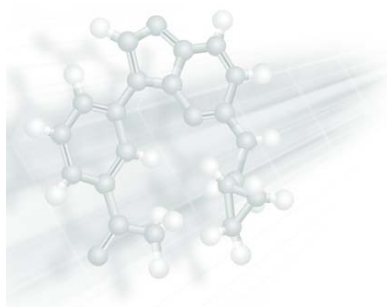
### How do you assess chemical similarity?

The chemical similarity of compounds can be based on molecular structure or any measured or predicted properties of the molecules, or a combination of the two. The most common approach is to use the molecular structure to assess the chemical similarity within a set of compounds. This approach defines the similarity in terms of the patterns of atoms present in the structures. The patterns of atoms along 'paths' through the 2D chemical structure are encoded in a binary 'fingerprint' and the similarity of two compounds defined in terms of the Tanimoto index. The advantage of a path-based fingerprint approach to similarity and diversity is that it provides a 'generic' method of comparing compounds. No assumption is made regarding the characteristics of molecules that will correlate most strongly with the biological activities of interest. Also, similarity assessed in this way usually corresponds well to a chemist's view; compounds from within a chemical series will typically have a high Tanimoto index.

### What are the axes of the 'chemical structure' space plot?

It is not possible to quantify the axes of the structural chemical space plot as they do not correspond to a particular property or descriptor. The structural chemical space plot is essentially a two-dimensional approximation of a multi-dimensional space, where similarity between each molecule in the data set represents a single dimension in the multi-dimensional space. Using an approximate form of Principle Component Analysis (PCA) (full PCA would be prohibitively slow for large numbers of molecules), this is reduced to a two-dimensional space that maximizes the visible variation. As a result, the two dimensions are effectively a function of the molecules' similarities and do not have an explicit meaning.

The dimensions are best considered as distance measures such that the closer molecules are in the plot, the greater their similarity. Bear in mind that the more molecules there are in the plot, the more approximate the visible distance will be for any two individual points, although the overall trend covers the diversity of the entire set.



## FAQ

### StarDrop: *Chemical Space*

#### **How is the 'property space' plot created?**

Property chemical space plots employ a simple PCA approach and enable visualization of the distribution of compounds with respect to their properties. Groups of compounds with a similar profile of properties will cluster together in this type of space. Additionally, if specific descriptors are expected to correlate with an important property, these may be imported into StarDrop and used to define a chemical space. Thus, the diversity of selections with respect to these descriptors may be visualized to aid in library design.

#### **Can you define chemical space using your own descriptors?**

Yes, StarDrop can build chemical space projections using any combination of continuous, categorical or structure data, including imported data.

#### **What is the benefit of a 'mixed' structure/property space plot over the other two options?**

Space plots defined using both chemical similarity and user-selected properties tend to better separate different chemotypes if they exhibit a similar spectrum of properties.

#### **How many compounds can I use in a chemical space plot?**

The chemical space display has no hard limits. However, generating the underlying chemical space plot will take a prohibitively long time for more than ~10,000 molecules. For sets bigger than this, it is worth using the Selection tool to perform a random selection of up to 10,000 molecules and use this selection to generate the chemical space. It is then possible to project any number of molecules into this space. However, the chemical space plot has been designed to work most efficiently for displaying sets of up to about 100,000 molecules on typical desktop hardware.

#### **Can you run Chemical Space without a server?**

Yes, this runs on the local machine.