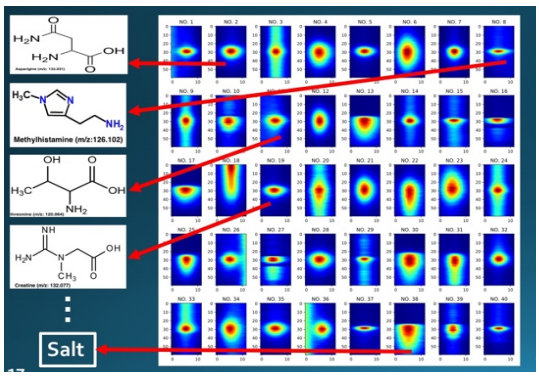
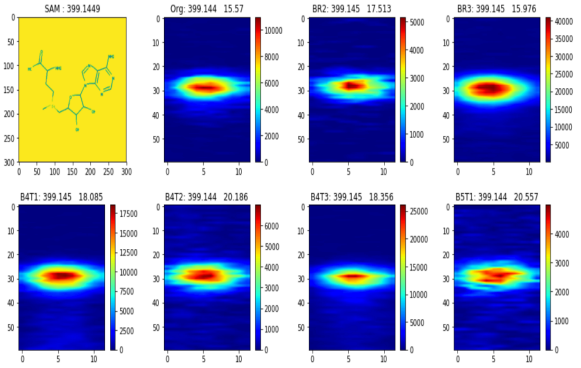
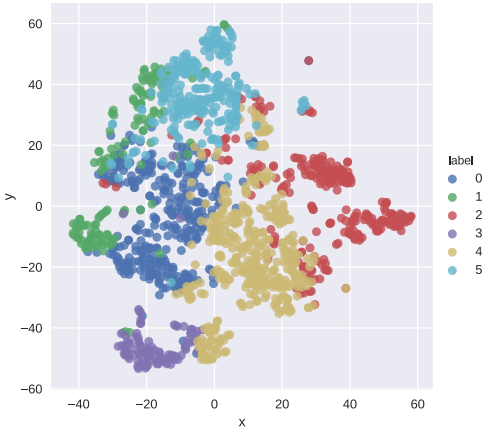
Project 2. Relation: signal shape & chemical properties?

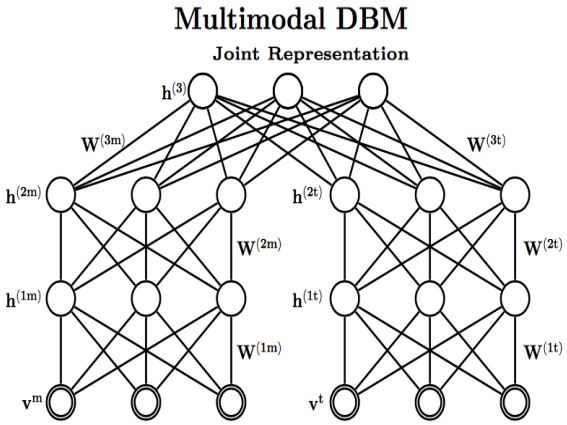
In the first project, we have trained our machine to detect the good signals from MS data from a single cell. It is found that even these good signals come with different shapes. It is noticed that some signals show very unique shape and consistent for that signal from different cell samples. Below shows one such example: *S-adenosyl-methionine* (SAM) from 7 samples show very similar signal shape. Is there any relation between the signal shape of the metabolites and their chemical properties?





Extended Connectivity Fingerprints (ECFPs) are circular topological fingerprints designed for molecular characterization. Here we use ECFP as the representation of chemical properties and the ECFPs (128 bits vector) of 32 different signals will be provided, as well as the 32 signal shape images (60*12, same as project 1) from 7 cell samples.





You are expected to:

1. Conduct some clustering (k-means, PCA, etc) and visualization of both the images and ECFPs to see: (1) If the same signal from different samples are very similar, or say, whether they are clustered into the same group? (2) If the clustering of the images and ECFPs match with each other, i.e., if signals are very similar in image shape, whether they are also very close for ECFPs?

2. Explore ML/DL models to decipher the relation between them. Multimodal learning with DBM may be a good choice for this task. As test, ECFPs of several new signals are to be predicted given their signal shape. (http://papers.nips.cc/paper/4683-multimodal-learning-with-deep-boltzmann-machines)