Machine Learning for Ionic Conductivity

Machine Learning for Ionic Conductivity

Topics: [Energy Capture & Storage](), [Informatics and Team Collaboration]()

Methodology: [Machine Learning]()

Products Used: [MS Maestro](), [AutoQSAR](), [MS Informatics]()

This tutorial is written for use with a 3-button mouse with a scroll wheel.

Words found in the **Glossary of Terms** are shown like this: Workspace

Abstract:

In this tutorial, we will learn how to develop machine learning models to predict the experimental ionic conductivity of ionic liquids.

Tutorial Content

1. Introduction

Quantitative Structure-Activity Relationships (QSAR) are useful modeling tools to efficiently predict material properties for a wide-range of molecules. Schrödinger's AutoQSAR tools for generating machine learning models are easy to use, facilitating automated generation of accurate QSAR models. For practice, tutorials are available using the Materials Science (MS) Maestro suite to predict properties of small molecules, polymers, and periodic systems: Machine Learning for Materials Science, Polymer Descriptors for Machine Learning, Cheminformatics Machine Learning for Homogeneous Catalysis and Periodic Descriptors for Inorganic Solids.

The current generation of Li-ion batteries use carbonate-based electrolytes mixed with salts, such as lithium hexafluorophosphate, $LiPF_6$, which are highly volatile and flammable. To address safety concerns arising from hazardous electrolytes, one possible solution is to replace the electrolyte with an ionic liquid (IL). Ionic liquids have good electrochemical and thermal stability, which could result in a safer battery; however, they suffer from low-to-medium ionic conductivity, which dictates how fast a battery can charge or discharge. Significant efforts have been focused on identifying ILs that have high ionic conductivities while maintaining the stability gained from using these electrolytes. In this tutorial, we will use the AutoQSAR panel in MS Maestro and an IL dataset from the NIST IL Thermo Database to create a machine learning model to predict ionic conductivity on a set of ILs at a fixed temperature of ~298.15 Kelvin (see **References**). A total of ~400 ILs are used to train and evaluate the machine learning models.
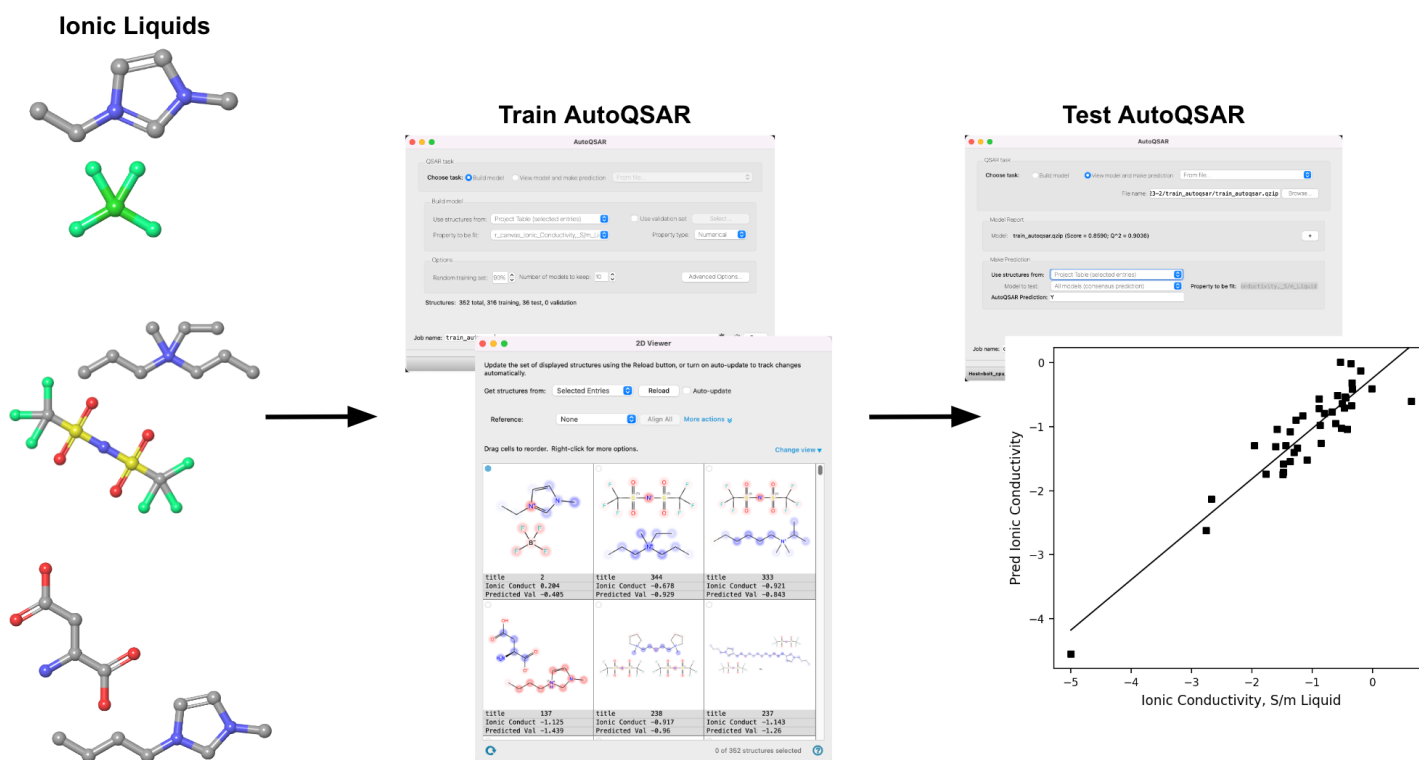

Figure 1. Tutorial workflow showing the 3D structures of ionic liquids in MS Maestro, which are used to train and test machine learning models with the AutoQSAR panel.

2. Creating Projects and Importing Structures

At the start of the session, change the file path to your chosen Working Directory in MS Maestro to make file navigation easier. Each session in MS Maestro begins with a default Scratch Project, which is not saved. A MS Maestro project stores all your data and has a `.prj` extension. A project may contain numerous entries corresponding to imported structures, as well as the output of modeling-related tasks. Once a project is saved, the project is automatically saved each time a change is made.

Structures can be built in MS Maestro or can be imported using **File > Import Structures** (or drag-and-dropped), and are added to the Entry List and Project Table. The Entry List is located to the left of the Workspace. The Project Table can be accessed by **Ctrl+T (Cmd+T)** or **Window > Project Table** if you would like to see an expanded view of your project data.
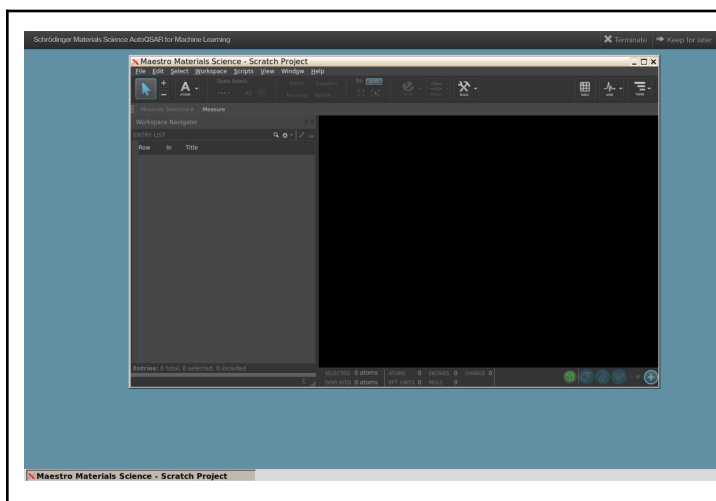
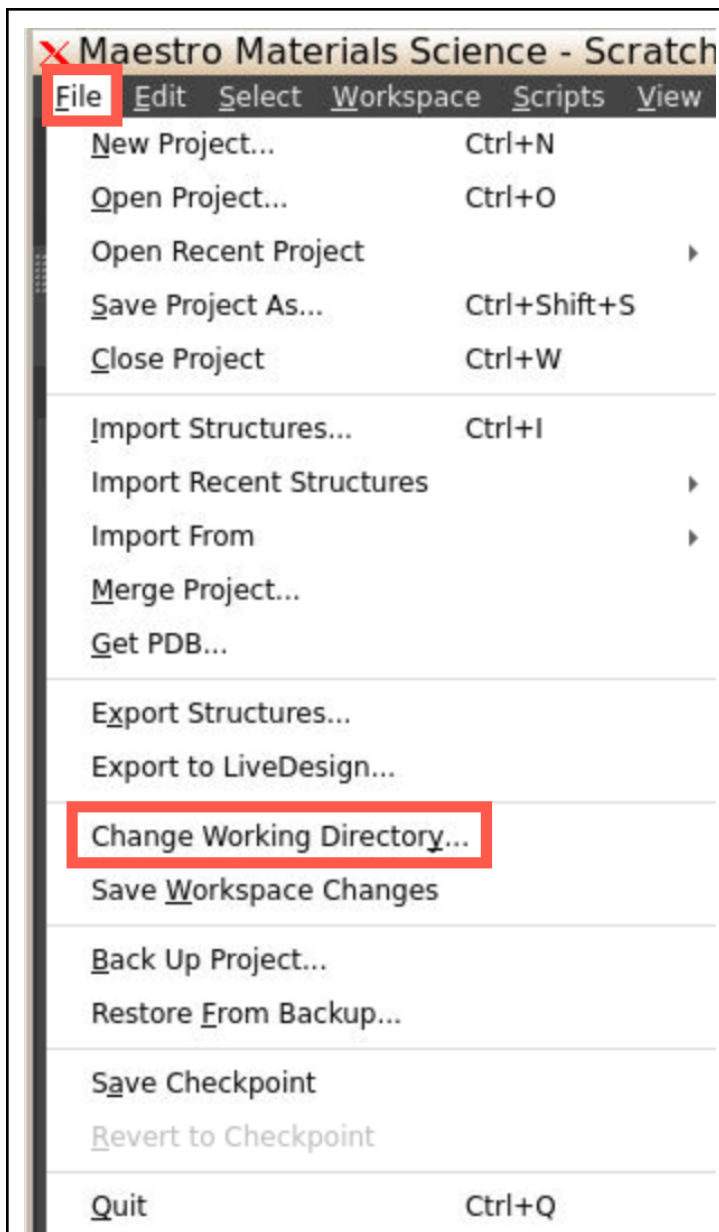|	
	1. Launch the tool in the nanoHUB interface ○ Launching the tool will automatically open up MS Maestro

Figure 2-1. Change Working Directory option.
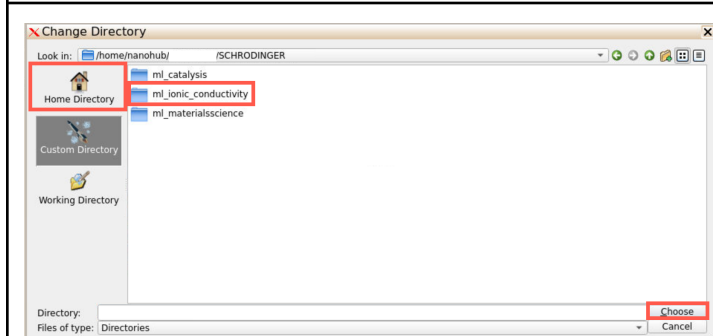
2. Go to **File > Change Working Directory**

Figure 2-2. Selecting the Working Directory.

3. Navigate to your Home Directory then the SCHRODINGER directory
4. Select *ml_ionic_conductivity*, and click **Choose**
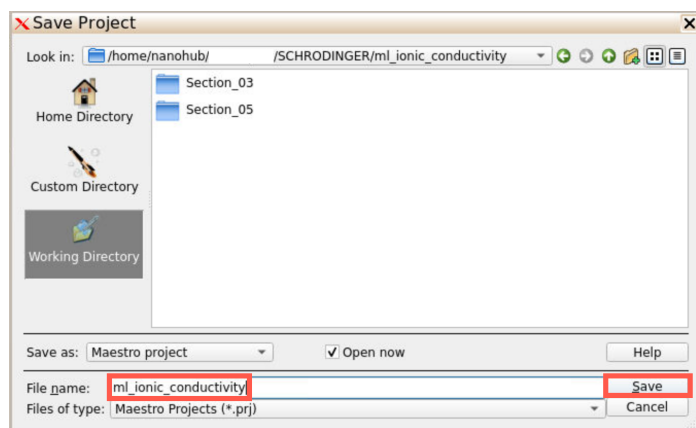   ○ All files needed to execute this tutorial are included in this directory

Figure 2-3. Save Project panel.

5. Go to **File > Save Project As**
6. Change the *File name* to **ml_ionic_conductivity**, click **Save**
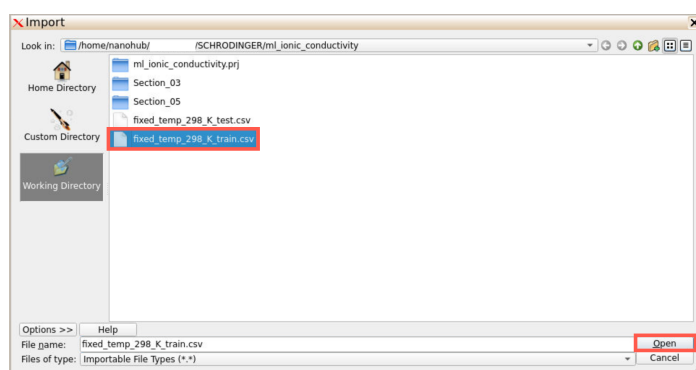   ○ The project is now named `ml_ionic_conductivity.prj`

Figure 2-4. Selecting the training file.

7. Go to **File > Import Structures**
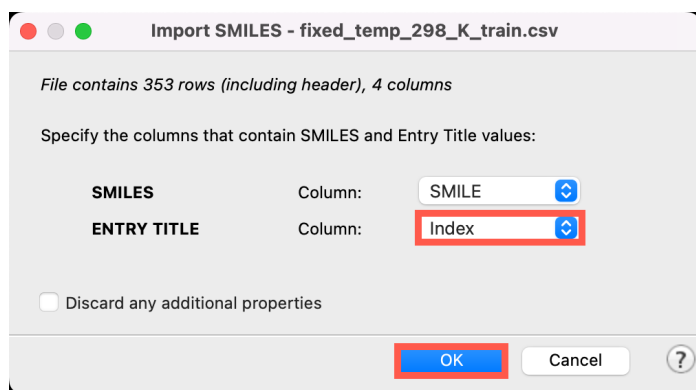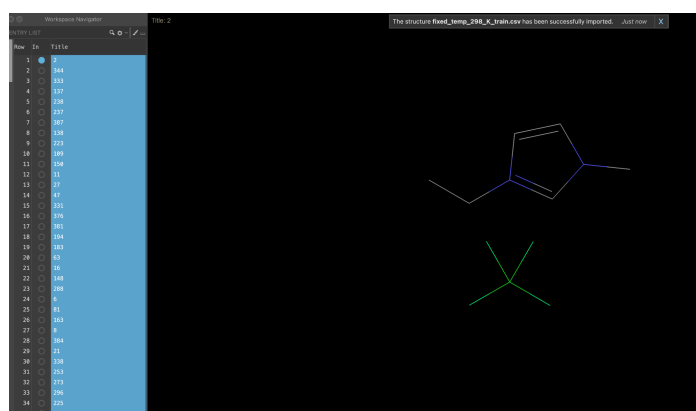8. Navigate to your working directory and choose `fixed_temp_298_K_train.csv`. Click **Open**

Figure 2-5. Importing in the training set.

9. Change the *ENTRY TITLE Column* to **Index**
10. Click **OK**

352 ionic liquid structures with their respective ionic conductivity values that will be used to train AutoQSAR models have now been imported into the project

Note: The ionic conductivity values are associated with each entry, and can be viewed in the Project Table. Use the Property Tree to expose All > Canvas > Secondary > Ionic Conductivity, S/m Liquid

| *FIgure 2-6. The ionic conductivity data set has now been imported.* | |

3. Building Machine Learning Models Using AutoQSAR

In this section, we will use the AutoQSAR panel to build machine learning models on ionic conductivity for a group of ionic liquids.

Figure 3-1. Grouping the training set.

1. In the entry list, <u>select</u> all the entries and right click
   - Holding the shift key will allow you to select all entries
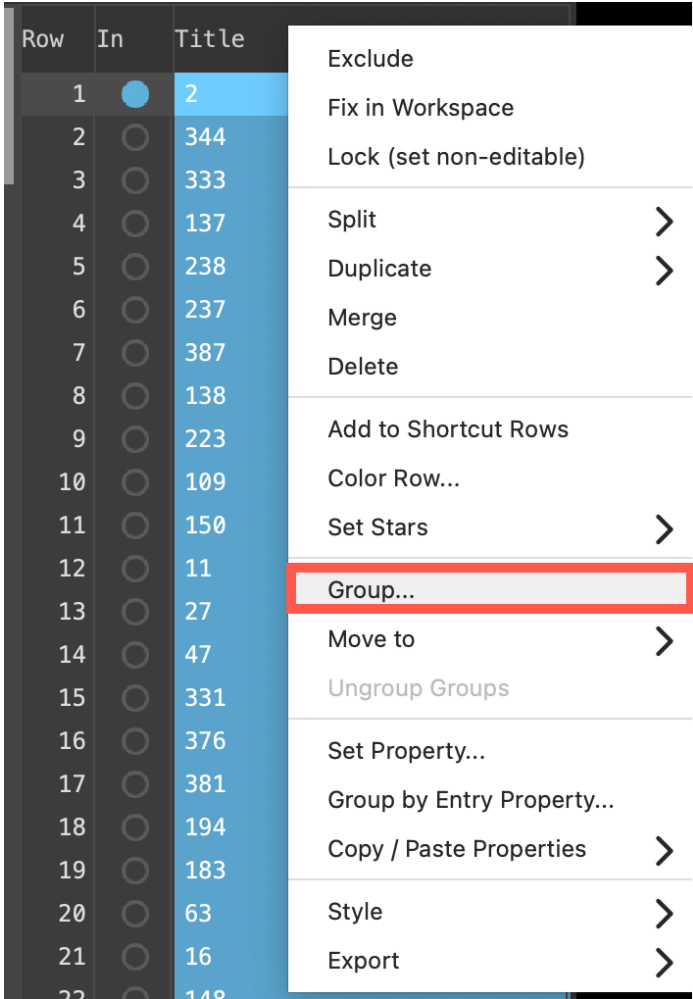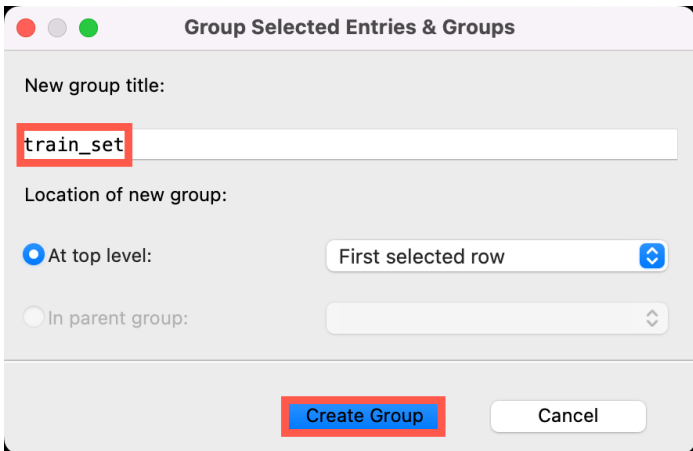2. Click **Group** to create a single entry group for the training set

Figure 3-2. Naming the new group.

3. For *New group title* input **train_set**
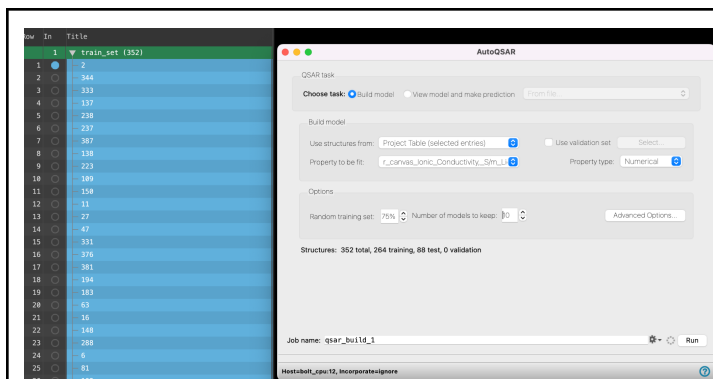4. Click **Create Group** to close the panel

Figure 3-3. Opening the AutoQSAR panel.

5. In the entry list, <u>select</u> the **train_set (352)** group
6. Go to **Tasks > Materials > Informatics > AutoQSAR**
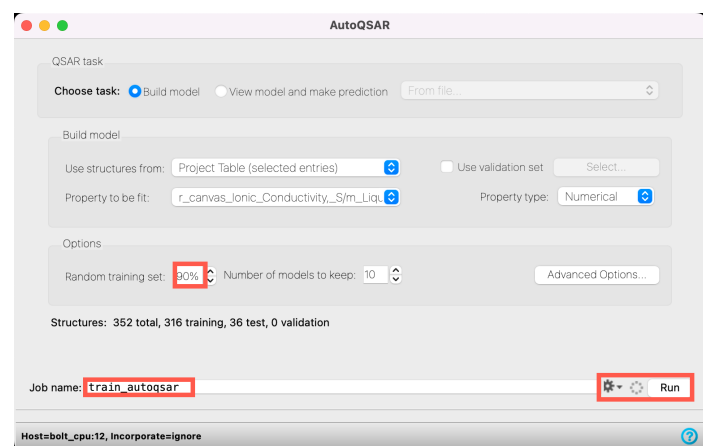   - The [AutoQSAR](#) panel opens

Figure 3-4. Setting the random training split.

7. Set **90%** for the *Random training split*
   - This is the percentage of data to set aside between train and test sets, where 90% of the data is used to train the model and 10% of the data is used to test the model
8. Maintain the remaining parameters:
   - The *Property to be fit* is the ionic conductivity data
   - The *Property type* is numerical (as opposed to categorical)

More options for defining models are available in the *Advanced Options* dialog box, and can be referenced in the [help documentation](#).

9. Change the *Job name* to **train_autoqsar**
10. Adjust the job settings (⚙) as needed
    - This job requires a CPU host and can be completed in about 30 minutes
11. If you would like to run the job yourself, click **Run**. Otherwise, use the pre-generated `Section_03 > train_autoqsar > train_autoqsar.qzip` file from the provided tutorial files in the following section
12. **Close** the AutoQSAR panel

4. Viewing the Machine Learning Models

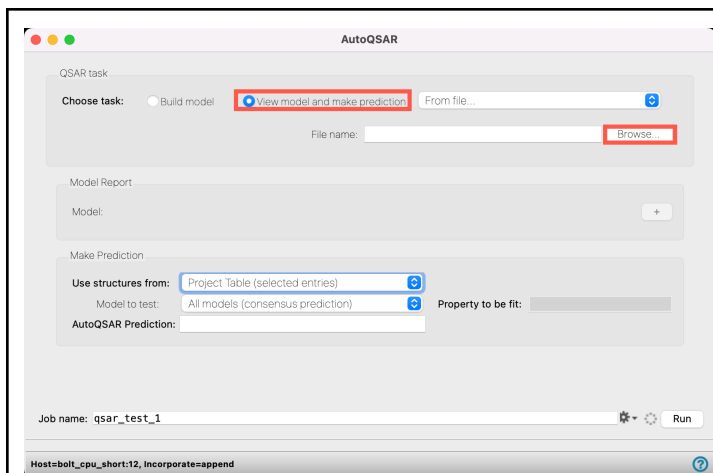Using the AutoQSAR panel, we can analyze the generated models.

Figure 4-1. Opening the AutoQSAR panel.

When the job is complete, note that no new entry group is added to the entry list. The output can be analyzed and used for predictions in the AutoQSAR panel:

1. Return to **Tasks > Materials > Informatics > AutoQSAR**
   - The AutoQSAR panel opens
2. For *Choose task*, switch to **View model and make prediction**
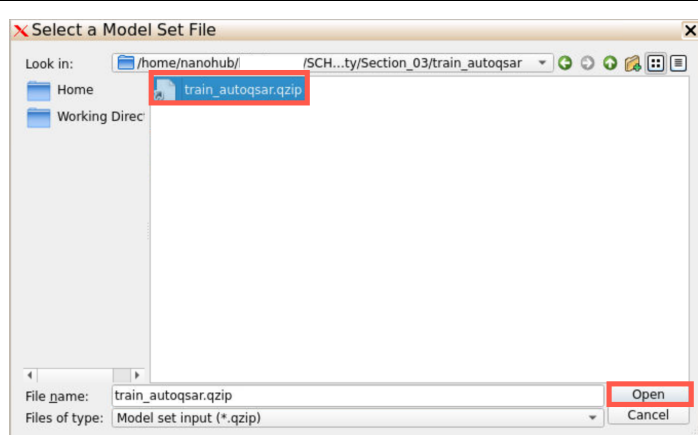3. To choose the *Model file* click **Browse**

Figure 4-2. Loading the .qzip file.

4. If using the provided file, navigate to the `Section_03 > train_autoqsar > train_autoqsar.qzip` file and click **Open**
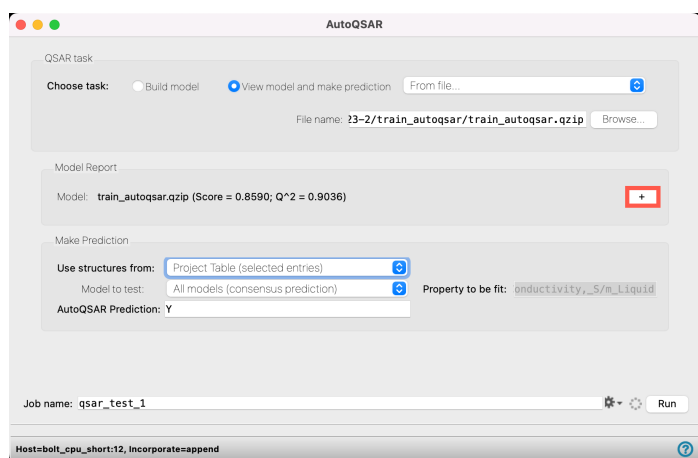   - The panel will parse the .qzip file and the *Model Summary* section will be populated

Figure 4-3. Expanding the model report.

5. In the *Model Report* section, **click the + button**
   - The *Model Report* section of the panel shows the ranking score and Q^2 value (the R^2 for the test set) for the best models

Several hundred 2D descriptors (molecular, topological, and feature counts) are automatically generated during the model building process and the top 10 models are retained. The models retained have performed well on the test set and have highly consistent training and test set statistics which results in a

higher ranking score.

The report gives the score and Q^2 value of the best-ranked models from the set of models generated in the run.

Figure 4-4. Visualizing the models.

The informative descriptors and fingerprints are used to build a large number of numeric or categorical models, where a given model is trained against a particular random subset of the input structures. The model is applied to the remaining input structures, and the accuracy of those predictions is used to arrive at an optimal number of factors for KPLS, PCR, and PLS models, and to assign an overall ranking score to the model.

In this case, the highest scoring models are all KPLS (Kernel partial least squares regression). For KPLS models, we can visualize atomic contributions to help understand the structural relationship to the model.

For each structure, each atom that contributed to a fingerprint used in building the model can be marked with a colored disk that represents the value of the contribution to the property due to that atom. The disks are blue for negative values and red for positive values. The color saturation indicates the magnitude of the contribution. Atoms that did not appear in any fingerprint are not marked with a disk.
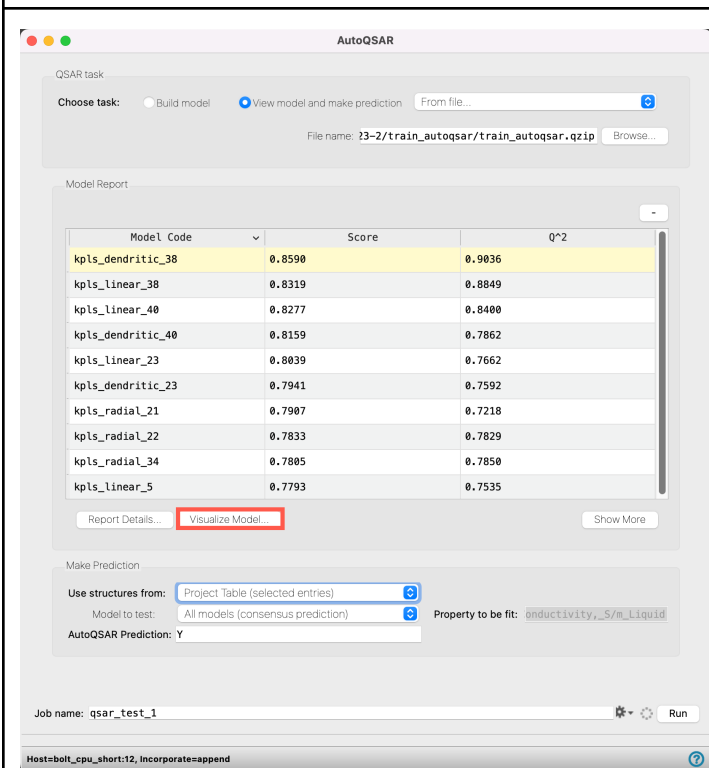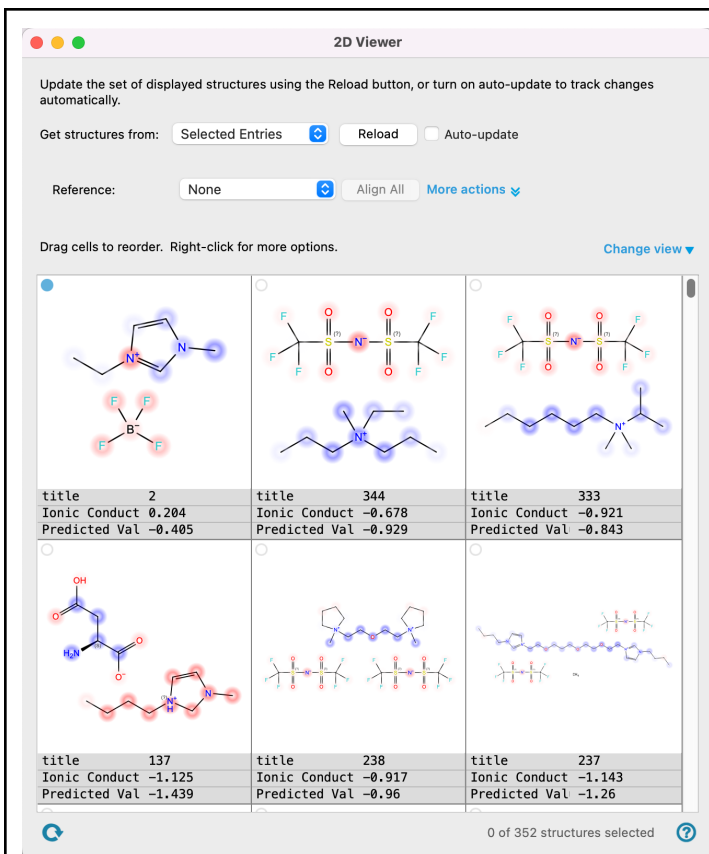
6. Click **Visualize Model**

Figure 4-5. Viewing the conductivity.

A 2D Viewer opens in which the blue regions contribute negatively to the ionic conductivity, whereas red regions contribute positively to the ionic conductivity. Such analysis may be useful for the future design of ionic liquids.

Feel free to explore the various structures in the viewer.

Close the 2D Viewer panel once finished.

Figure 4-6. Opening the Report.

7. Click **Report Details** to view the detailed report for the kpls_dendritic_38 model
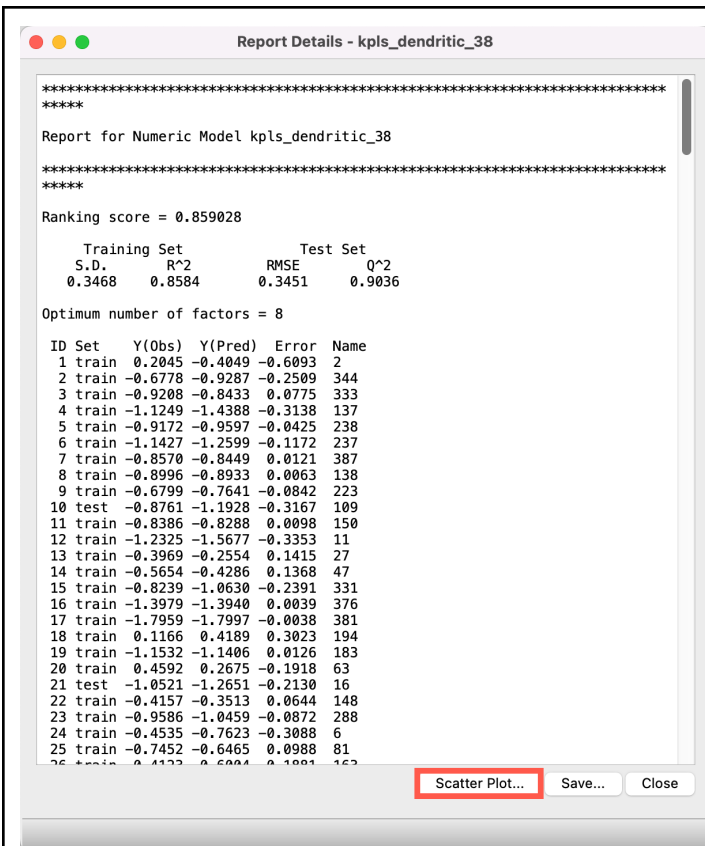
Figure 4-7. Viewing the Report Details.

The Report Details includes the ranking scores as well as the statistical values associated with the training and test sets. In addition, each IL and its observed versus predicted value can be found here.

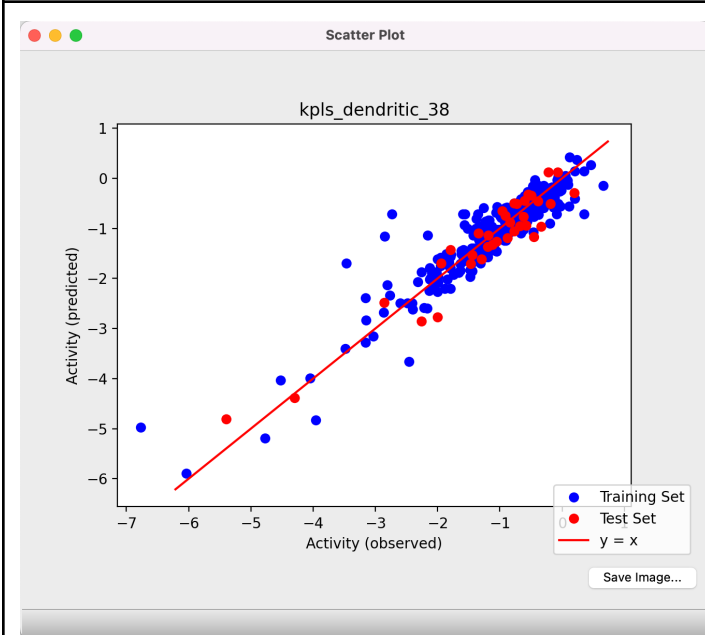8. To plot the data click **Scatter Plot**

Figure 4-8. Viewing the scatter plot.

This scatter plot visualizes the parity plot between predicted and actual ionic conductivity

9. **Close** the Scatter Plot, Report Details and AutoQSAR panel when finished

5. Predicting Ionic Conductivity for an Unseen Test Set

In this section, we will use the [AutoQSAR](#) panel to predict ionic conductivity on an unseen test data set of ionic liquids that were not used to train and evaluate AutoQSAR models.
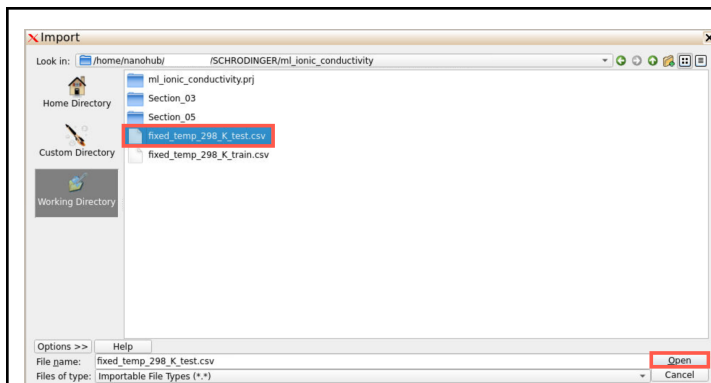

Figure 5-1. Selecting the test set file.

Let's import the test data set:

1. Go to **File > Import Structures**
2. Navigate to where you downloaded the tutorial files (presumably your working directory) and choose `fixed_temp_298_K_test.csv`. Click **Open**
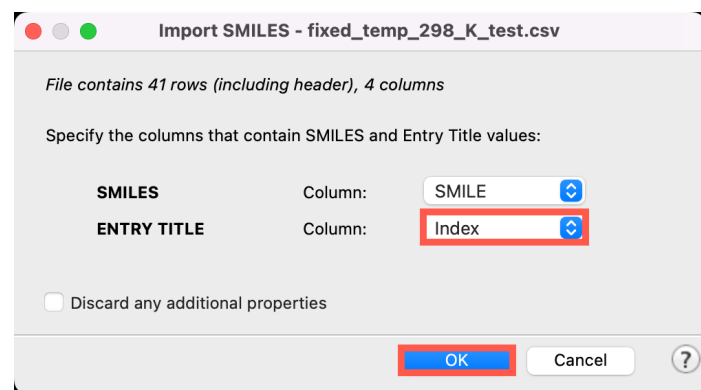

Figure 5-2. Importing in the test set.

3. Change the *ENTRY TITLE Column* to **Index**
4. Click **OK**
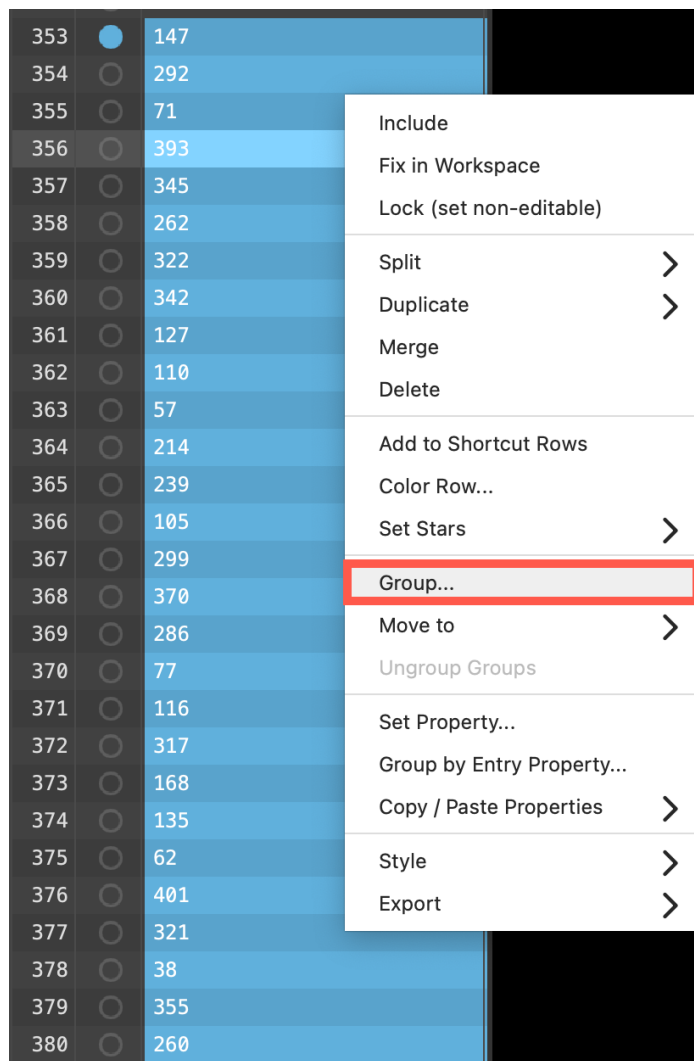
Figure 5-3. Grouping the test set.

A test set of 40 ionic liquid structures and ionic conductivity values have now been imported into the project. These structures were not used in the training and evaluation of the AutoQSAR models.

5. In the entry list, <u>select</u> all the newly imported test set entries and right-click
6. Click **Group** to group the test set

Figure 5-4. Naming the new group.

7. For *New group title* input **test_set**
8. Click **Create Group** to close the panel

Figure 5-3. Choosing the QSAR task.

9. In the entry list, <u>select</u> the test_set group
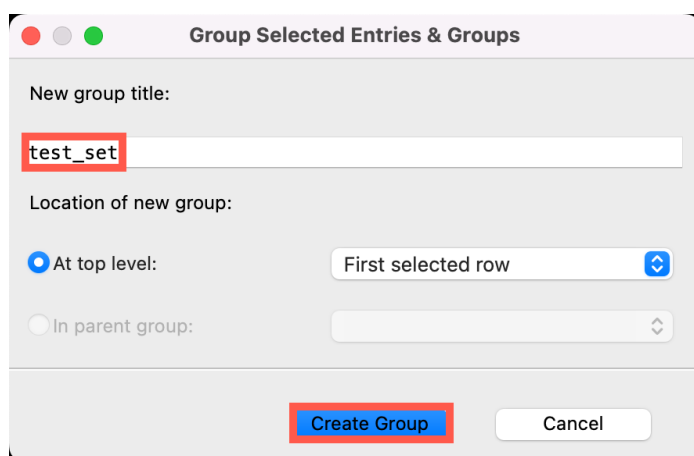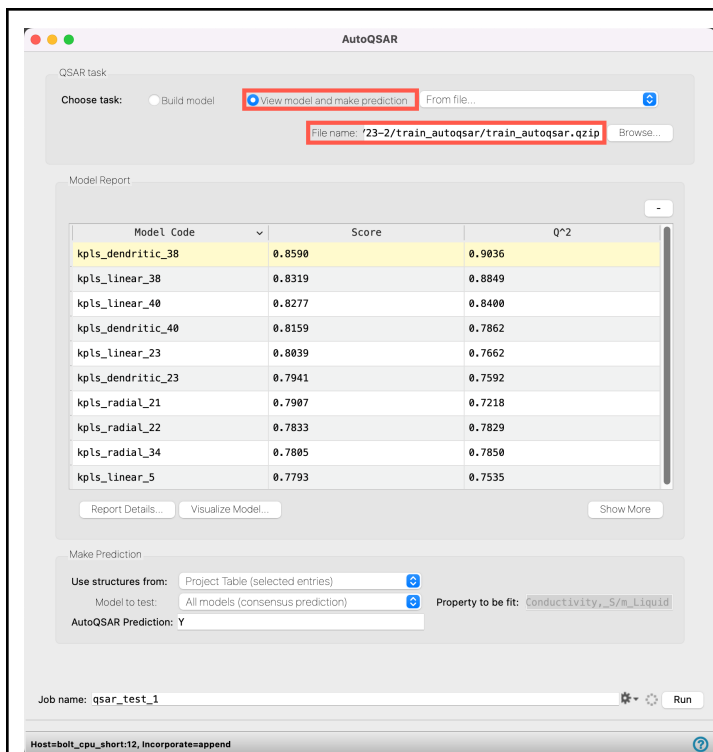10. Go to **Tasks > Materials > Informatics > AutoQSAR**
    - The AutoQSAR panel opens
    - In addition to using the AutoQSAR panel to build and evaluate our models, we can also use this panel to make predictions
11. For *Choose task* select **View model and make prediction**
12. Ensure that the `train_autoqsar.qzip` file is still selected
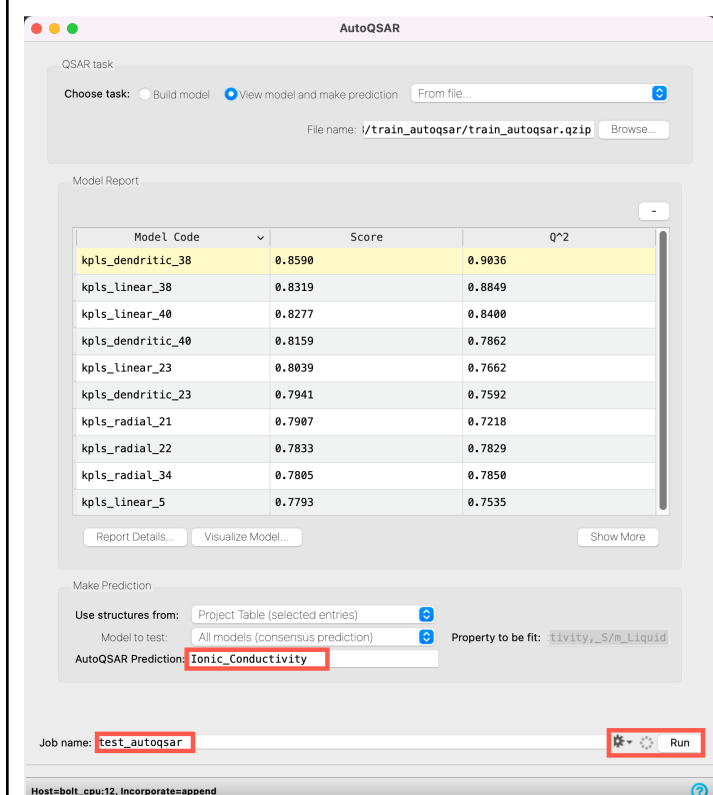

Figure 5-4. Running the job.

13. Change the *AutoQSAR Prediction* to **Ionic_Conductivity**
    - This will be name of the predicted values property

Note: Model to test is set to **All models (consensus prediction).** Consensus prediction averages the results of the retained models, which can often increase the accuracy of the predictions.

14. Change the *Job name* to **test_autoqsar**

15. Adjust the job settings ( ) as needed
    - This job requires a CPU host and can be completed in about 10 minutes
16. If you would like to run the job yourself, click **Run**. Otherwise, import the pre-generated `Section_05 > test_autoqsar > test_autoqsar-out.mae.gz` file from the provided tutorial files via **File > Import Structures**

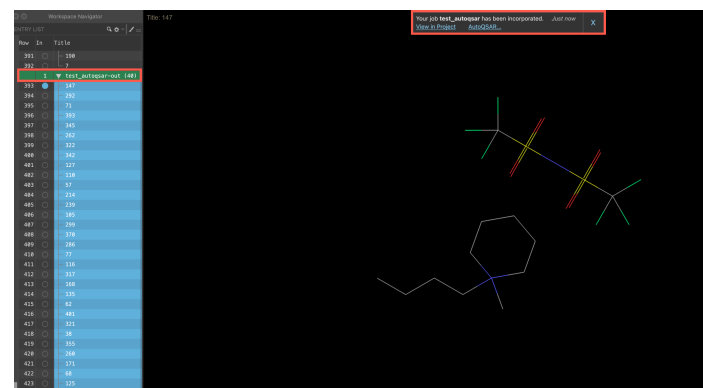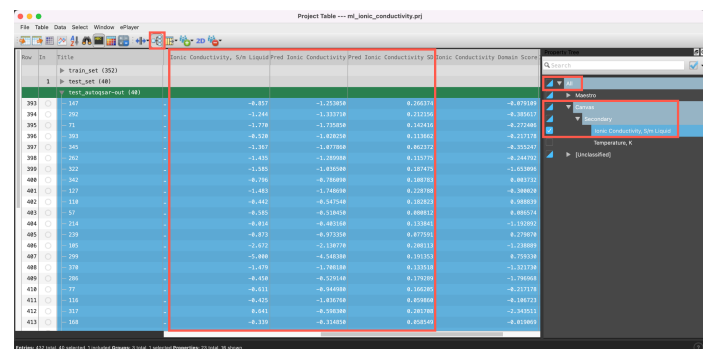|	
	17. **Close** the AutoQSAR panel
Figure 5-5. Viewing the output of the prediction task.	When the job is complete or after importing, a new entry group is added to the entry list titled **test_autoqsar-out (40)** containing the same 40 entries, but now with predicted ionic conductivity properties.
Figure 5-6. The Project Table.	The resulting data can be analyzed in the Project Table 18. Open the Project Table () 19. Use the Property Tree () to include the *Ionic Conductivity score* (**Check** the properties of interest under **All > Canvas > Secondary > Ionic Conductivity, S/m Liquid**) The Project Table should show "Ionic Conductivity" and predictions by AutoQSAR ("Pred Ionic Conductivity"). Approximate prediction uncertainties and domain scores from AutoQSAR are also shown as "Pred Ionic Conductivity SD" and "Predicted Ionic Conductivity Domain Score", respectively (high magnitudes of domain scores tell you whether a molecule is extremely distinct from the original training set).
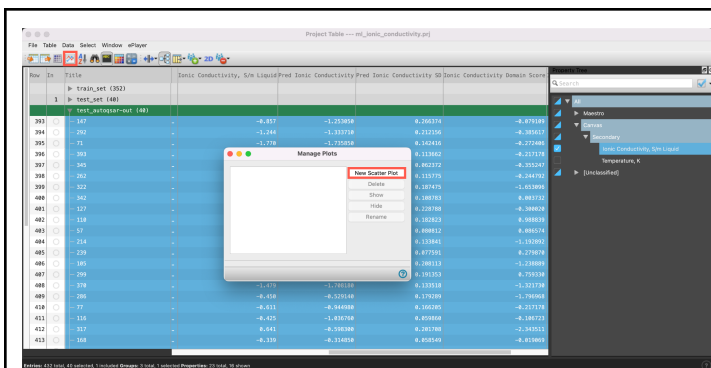
Figure 5-7. Creating a scatter plot.

20. Click the **Manage Plots** button ( )
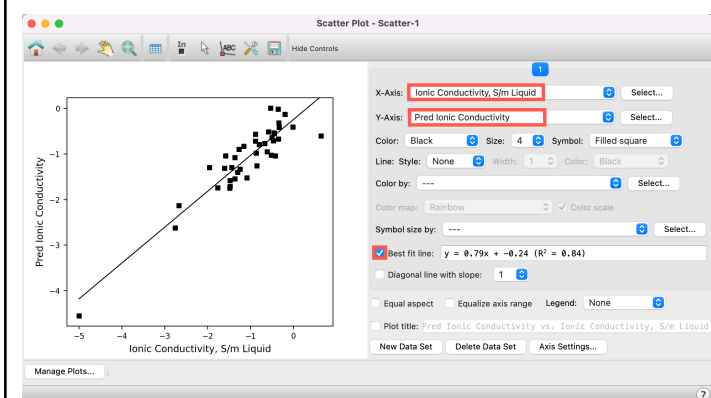21. Click **New Scatter Plot**

Figure 5-8. Plotting the ionic conductivity.

22. In the new scatter plot, change the following parameters:
    - X-axis: **Ionic Conductivity, S/m Liquid**
        - i. This is the experimental ionic conductivity
    - Y-axis: **Pred Ionic Conductivity**
        - i. This is the predicted ionic conductivity
    - Check **Best fit line**

The best fit line between predicted and actual values shows a reasonable R^2 of 0.84 (an ideal model would have an R^2 of 1.00). The results suggest that the ML model derived could be used to predict ionic conductivity values for ionic liquids. Furthermore, this workflow highlights the computational efficiency achieved when using ML approaches as compared to other computational (*e.g. ab initio* calculations) or experimental approaches. While this tutorial uses a relatively small dataset, one could envision a larger training set would further improve prediction accuracy.

6. Conclusion and References

In this tutorial, we learned how to use the [AutoQSAR panel](#) to build machine learning models to predict the ionic conductivity for ionic liquids. These machine learning models enable fast screening of ionic liquids for high ionic conductivities, which could help build safer and more efficient batteries.

For further learning:

For introductory content, focused on navigating the Schrödinger Materials Science interface, an Introduction to Maestro for Materials Science tutorial is available. Please visit the materials science training website for access to 50+ tutorials. For scientific inquiries or technical troubleshooting, submit a ticket to our Technical Support Scientists at help@schrödinger.com.

For self-paced, asynchronous, online courses in Materials Science modeling, including access to Schrödinger software, please visit the Schrödinger Online Learning portal on our website.

For some related practice, proceed to explore other relevant tutorials:

- For more machine learning practice with MS Maestro:
  - Machine Learning for Materials Science
  - Polymer Descriptors for Machine Learning
  - Periodic Descriptors for Inorganic Solids
  - Machine Learning Property Prediction
  - Optoelectronics Active Learning
  - Machine Learning for Sweetness
  - Cheminformatics Machine Learning for Homogeneous Catalysis

- For general battery-related workflows:
  - Liquid Electrolytes Properties: Part 1
  - Liquid Electrolyte Properties: Part 2
  - Calculating Voltage Curves of Spinel Intercalation Compounds
  - Polymer Electrolyte Analysis
  - Diffusion

For further reading:

- The dataset was extracted from the supplementary information of: Conductivity prediction model for ionic liquids using machine learning. DOI: 10.1063/5.0089568
- The original dataset is from: NIST IL Thermo Database
- Developing machine learning models for ionic conductivity of imidazolium-based ionic liquids. DOI: 10.1016/j.fluid.2021.113208
- A generalized machine learning model for predicting ionic conductivity of ionic liquids. Molecular Systems Design and Engineering. DOI: 10.1039/D2ME00046F
- DeepAutoQSAR: Scalable, Intuitive, Deep-learning QSAR models for Big Data Applications (Schrödinger white paper)
- DeepAutoQSAR Hardware Benchmark (Schrödinger white paper)
- See the help documentation for more information on the AutoQSAR panel

7. Glossary of Terms

<u>Entry List</u> - a simplified view of the Project Table that allows you to perform basic operations such as selection and inclusion

<u>Included</u> - the entry is represented in the Workspace, the circle in the In column is blue

<u>Project Table</u> - displays the contents of a project and is also an interface for performing operations on selected entries, viewing properties, and organizing structures and data

<u>Recent actions</u> - This is a list of your recent actions, which you can use to reopen a panel, displayed below the Browse row. (Right-click to delete.)

<u>Scratch Project</u> - a temporary project in which work is not saved, closing a scratch project removes all current work and begins a new scratch project

<u>Selected</u> - (1) the atoms are chosen in the Workspace. These atoms are referred to as "the selection" or "the atom selection". Workspace operations are performed on the selected atoms. (2) The entry is chosen in the Entry List (and Project Table) and the row for the entry is highlighted. Project operations are performed on all selected entries

<u>Working Directory</u> - the location where files are saved

<u>Workspace</u> - the 3D display area in the center of the main window, where molecular structures are displayed