

AutoSherlockTM



"Because scientists would rather spend less time organizing their data, and more time learning from it."

Version 1.4.2

Ray Nagel and Edward Snell

Crime scene walkthrough

Note: This document covers the development version of AutoSherlock and is not intended for distribution beyond the initial testers yet.

NOTE: The current version of AutoSherlock is 1.16.2 which has features that are not included in this manual (update in progress).

TABLE OF CONTENTS

1. Introduction.....	3
2. Frequently asked questions (and sometimes answers).....	4
3. Requirements	6
4. Current Version	7
4.1 Updates from version 1.3	7
4.2 Updates from version 1.2	7
4.3 Updates from version 1.1	7
5. Installation	8
6. Operation.....	9
6.1 Input files and output files	9
6.2 AutoSherlock	9
7. The Sherlock spreadsheet.....	19
7.1 The Outcome Summary	19
7.2 The HWI and Hampton Research Chemical Space Summary.....	22
7.3 The Cocktail Listing results output	26
8. How to Use Sherlock.....	27
8.1 How results relate to each other.	27
8.2 Conditions that have not been sampled.....	28
8.3 Identify incorrect scores	28
8.4 Optimization	31
8.5 Other features	32
9. Known Bugs	34
10. Unknown Bugs	35
11. Acknowledgements and references.	36

1. Introduction

AutoSherlock is a program designed to present results from the 1536 screen in chemical space. It takes in scored or un-scored files from MacroScope and creates an Excel spreadsheet linked to the crystallization screening images. The spreadsheet can be used to visualize results in chemical space, check the accuracy of scoring and direct subsequent optimization.

2. Frequently asked questions (and sometimes answers)

It's taking a long time to do something, should I restart it?

AutoSherlock translates a large file with chemical data in, sorts through that file to reformulate it in chemical space and constructs an Excel file from scratch. Unfortunately, the process, while rapid to code, is not that efficient. For the first time running follow section 5 and be patient. There are a number of instances where the progress looks stalled, but the program is running. Go get a cup of coffee and then come back later! We apologize about this and plans are underway to write a standalone version that does not have these problems. It does run slightly faster if the Excel file sheet is not shown.

I can't see the crystallization images when I click on the spreadsheet?

There are a number of possible causes for this. The wrong directory may have been entered when first running Sherlock. However, the most common mistake we have seen is not allowing macros to be enabled when first opening the Sherlock spreadsheet. Symptoms of this are that only chemical information is seen when you click on a cell and no image appears.

When I send the Sherlock file to someone else they cannot see the images?

This has been a classic problem. Within the Sherlock spreadsheet there is only 1 cell that can be edited. This is cell A1 on the 'HWI Chemical Space' tab and this contains the directory name where the images are located. If you move the images make sure to alter this directory name to reflect the new image location.

The spreadsheet disappears after AutoSherlock finishes!

Annoying isn't it? This is actually a designed feature. Within the development laboratory we anticipate running AutoSherlock automatically on some 9000 different macromolecular samples in our archives if we figure out how to score all those images (and we are working on that). That would be a lot of spreadsheets open on one screen. In a future version we will put an option in to keep the spreadsheet open but current development is focused on a version that does not use Excel.

Can I rescore images in the Sherlock spreadsheet?

No. This would be superb if you could. The only option is to return to MacroScope, put the new score in and then rerun AutoSherlock. The reason for this is how Excel handles conditional formatting (i.e. the pretty colors). It will only allow 3 colors and the way we have worked round this is to specifically specify

the colors as the spreadsheet is generated. To flog a dead horse, a new version, not relying on Excel, is under development.

I can't seem to score images in Macroscopic?

Two possibilities here, when you select an individual image a scoring option is not available from the actions menu on the top of the image. If this is not seen then you have an old version of Macroscopic and you need to get the latest release. Contact Ray Nagel (address at the end of the document).

Do you mean I have to score all the images myself?

No and yes. You only need to score as many images as you feel like. You'll get more from the results if they are all scored but AutoSherlock does not care one way or the other. For now, you do have to score them yourself. Partial to full automated image analysis is on the horizon but not available for release yet.

What do I do with the list of cocktails that could not be mapped?

There are a couple of cocktails that are duplicated and because of this are not mapped. Others that are not mapped may have been changed from different generations. We update our cocktails slightly every year and are waiting on the development of an in-house database to format all the generations within Sherlock. Until that is in place there will be cocktails that are not mapped. The list is intended as a warning that you should check these and not exclude them from the analysis. We know of the ones that are not mapped but if you send the list to the contacts at the end of the document, we'll either deal with the problem or remove the list option in the next release or automatically forward all the emails to our database developer!

My question has not been answered here?

Let us know. If you have followed the walkthrough and problems have occurred we'd like to know about it and we'll try and address the question here. We're interested in anything to make it more useful as we use it also.

3. Requirements

Unofficially:

- Windows XP (no other version of Windows have been tested)
- Net Framework 3.0 or later
- Microsoft Excel 2003 to display the resulting spreadsheets
- MS Office 2003 (v11) Primary Interop Assembly Redistributable (installer included with the application)

4. Current Version

The current version is a Beta test version 1.4.2

4.1 Updates from version 1.3

- A number of bugs have been addressed or identified.
- A change in text browser has been made for the MSO file. This change should be invisible to the user.

4.2 Updates from version 1.2

- Default sort order of the cations and anions has been changed to an approximate Hofmeister series.
- Another information sheet has been added to give the chemical information, score and image name for use in other databases.
- Small formatting changes have been made.
- A results reports option is now available to identify any conditions that were not mapped successfully.

4.3 Updates from version 1.1

- AutoSherlock now incorporates a progress bar.
- The Hampton Research Cocktails are now tabulated.
- Known problems being addressed.
- The Hampton Research Cryo Screen incorporated as part of the initial 984 cocktails is not yet tabulated.

5. Installation

Soon to be available on a test basis - see Ray Nagel.

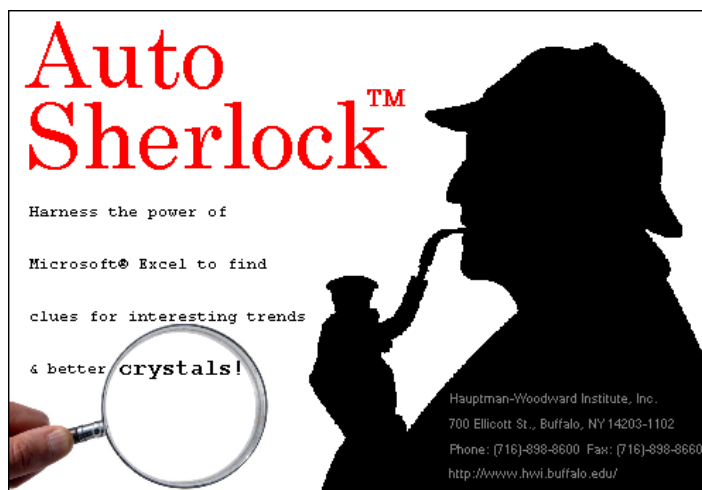
6. Operation

6.1 Input files and output files

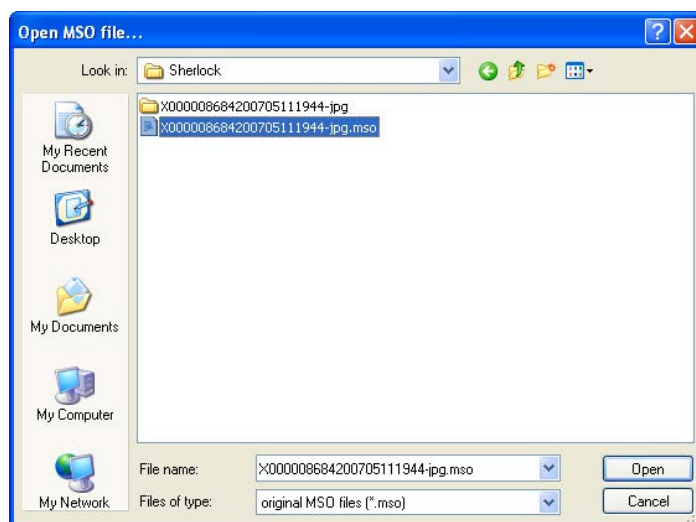
AutoSherlock operates on the output files generated from Macroscopic and requires uncompressed images from the rar files provided from the 1536 screening laboratory. A utility program called Backpacker can be used to generate these from the rar files. There are several other ways of obtaining these uncompressed files. To open them with Macroscopic and keep the Macroscopic program open or to quit the Macroscopic program when the files are open. This then requires locating where the files were uncompressed and is not an elegant solution. Macroscopic in the future will provide the option of saving the uncompressed files in a preferred location. The other method is to use the software, WinRar from <http://www.rarlab.com/>.

6.2 AutoSherlock

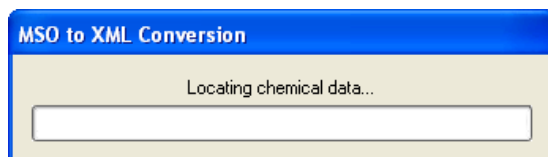
Double click on AutoSherlock to run. It will come up with the following window



Then ask you to choose an MSO file to open. The MSO files are text files that contain chemical information extracted by Macroscopic from the packaged image files. There are three file options, the default XML MSO file, the standard original MSO file and any file. We'll select an original MSO file to open where the file name is based on that from the imaging robot but could be changed to any name required. It is better to keep the original type description on the end.

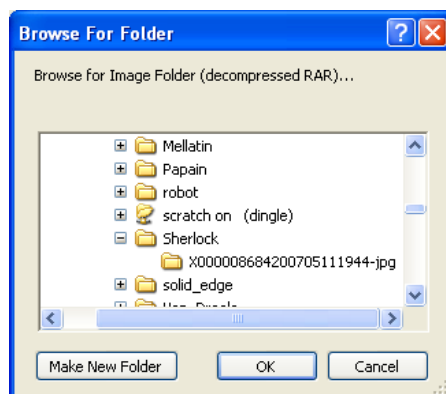


Another window, the main AutoSherlock window, also appears but is not active at this moment. Select the appropriate directory with your file in and hit open. In this case we will select the only file. A window opens telling us that it will automatically be converted to the new format. It first loads the chemical information from the file that is opened. This can take a minute or so and there will be no movement of the progress bar while it inputs the data.

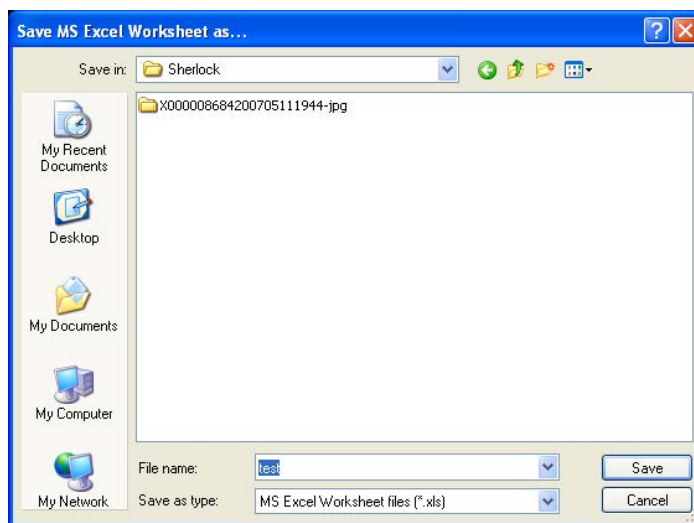


Once input AutoSherlock will write out the file into an XML readable format. A progress bar is shown while it does this.

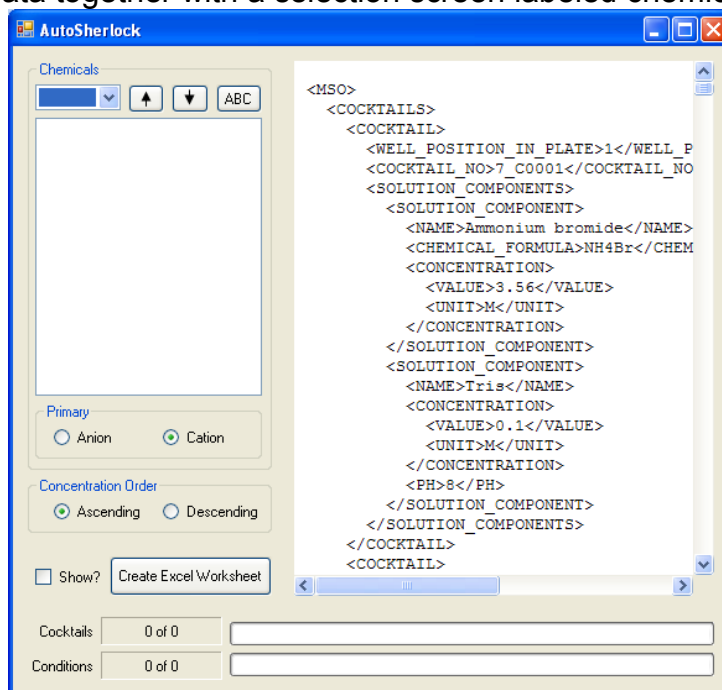
AutoSherlock asks you for the image folder containing the decompressed RAR file.



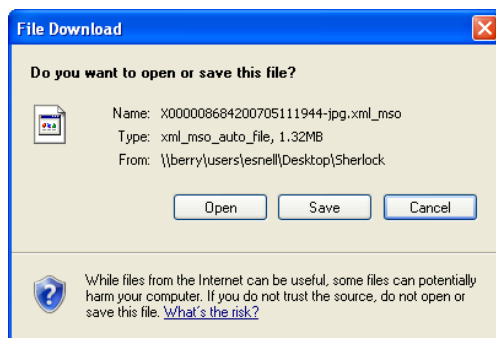
Select this and click OK. You will then be asked for the name and directory to put the output Spreadsheet. In this case, we have called it test. Hit save.



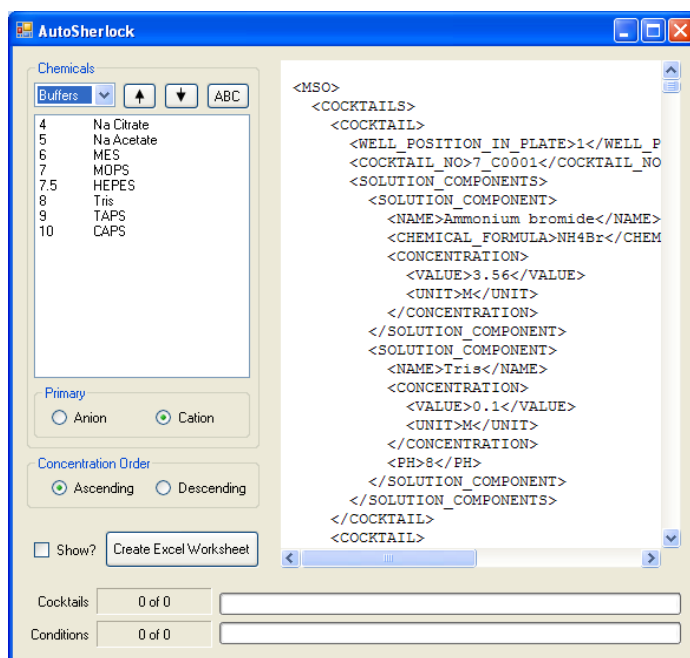
AutoSherlock will then present you with a screen showing the XML file containing the chemical data together with a selection screen labeled chemicals.



Note: There is a known bug where the XML file is not displayed. This is currently being addressed. If the program will not display the file and you are prompted to open, save or cancel, press cancel and then continue. You will not see the file information in the right hand window of AutoSherlock but the program is working as normal.



If this known bug is not affecting applicable to your system you can scroll down the XML file to see all the components of the screen. To create the chemical space you must select the chemicals from the box in the upper left hand corner. Currently you can select on Buffers, Anions, Cations and PEG. We recommend starting with Buffer. Select this and you are presented with a list of buffers.



You can highlight an individual buffer and move it up or down the list. By default the buffers are ordered from low to high pH.

The Anions, Cations and PEG are sorted in a similar manner. The Anions and Cations are ordered in approximate Hofmeister series. They can be reordered in any order the user wants.

AutoSherlock

Chemicals
 Anions ▼ ▲ ▾ ABC

phosphate-monobasic
 phosphate-dibasic
 phosphate-tribasic
 carbonate
 sulfate
 sulfate-monohydrate
 sulfate-hexahydrate
 sulfate-heptahydrate
 molybdate-hydrate
 molybdate-dihydrate
 acetate
 chloride
 chloride-dihydrate
 chloride-hexahydrate
 nitrate
 nitrate-hexahydrate
 bromide

Primary
☐ Anion ☒ Cation

Concentration Order
☒ Ascending ☐ Descending

☐ Show?

Cocktails 0 of 0

Conditions 0 of 0

```
<MSO>
<COCKTAILS>
<COCKTAIL>
  <WELL_POSITION_IN_PLATE>1</WELL_P
  <COCKTAIL_NO>7_C0001</COCKTAIL_NO
  <SOLUTION_COMPONENTS>
    <SOLUTION_COMPONENT>
      <NAME>Ammonium bromide</NAME>
      <CHEMICAL_FORMULA>NH4Br</CHEM
      <CONCENTRATION>
        <VALUE>3.56</VALUE>
        <UNIT>M</UNIT>
      </CONCENTRATION>
    </SOLUTION_COMPONENT>
    <SOLUTION_COMPONENT>
      <NAME>Tris</NAME>
      <CONCENTRATION>
        <VALUE>0.1</VALUE>
        <UNIT>M</UNIT>
      </CONCENTRATION>
      <PH>8</PH>
    </SOLUTION_COMPONENT>
  </SOLUTION_COMPONENTS>
</COCKTAIL>
</COCKTAILS>
```

AutoSherlock

Chemicals
 Cations ▼ ▲ ▾ ABC

Ammonium
 Rubidium
 Potassium
 Sodium
 Lithium
 Calcium
 Magnesium
 Cobalt
 Manganese
 Zinc

Primary
☐ Anion ☒ Cation

Concentration Order
☒ Ascending ☐ Descending

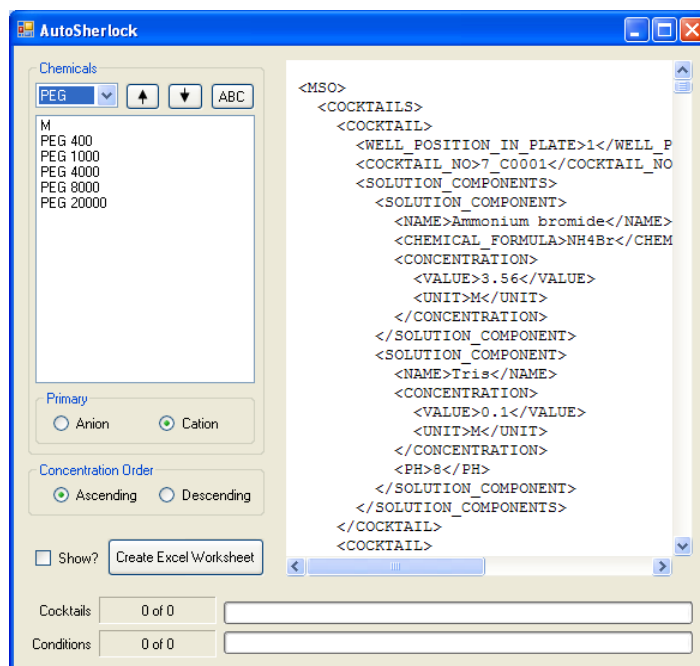
☐ Show?

Cocktails 0 of 0

Conditions 0 of 0

```
<MSO>
<COCKTAILS>
<COCKTAIL>
  <WELL_POSITION_IN_PLATE>1</WELL_P
  <COCKTAIL_NO>7_C0001</COCKTAIL_NO
  <SOLUTION_COMPONENTS>
    <SOLUTION_COMPONENT>
      <NAME>Ammonium bromide</NAME>
      <CHEMICAL_FORMULA>NH4Br</CHEM
      <CONCENTRATION>
        <VALUE>3.56</VALUE>
        <UNIT>M</UNIT>
      </CONCENTRATION>
    </SOLUTION_COMPONENT>
    <SOLUTION_COMPONENT>
      <NAME>Tris</NAME>
      <CONCENTRATION>
        <VALUE>0.1</VALUE>
        <UNIT>M</UNIT>
      </CONCENTRATION>
      <PH>8</PH>
    </SOLUTION_COMPONENT>
  </SOLUTION_COMPONENTS>
</COCKTAIL>
</COCKTAILS>
```

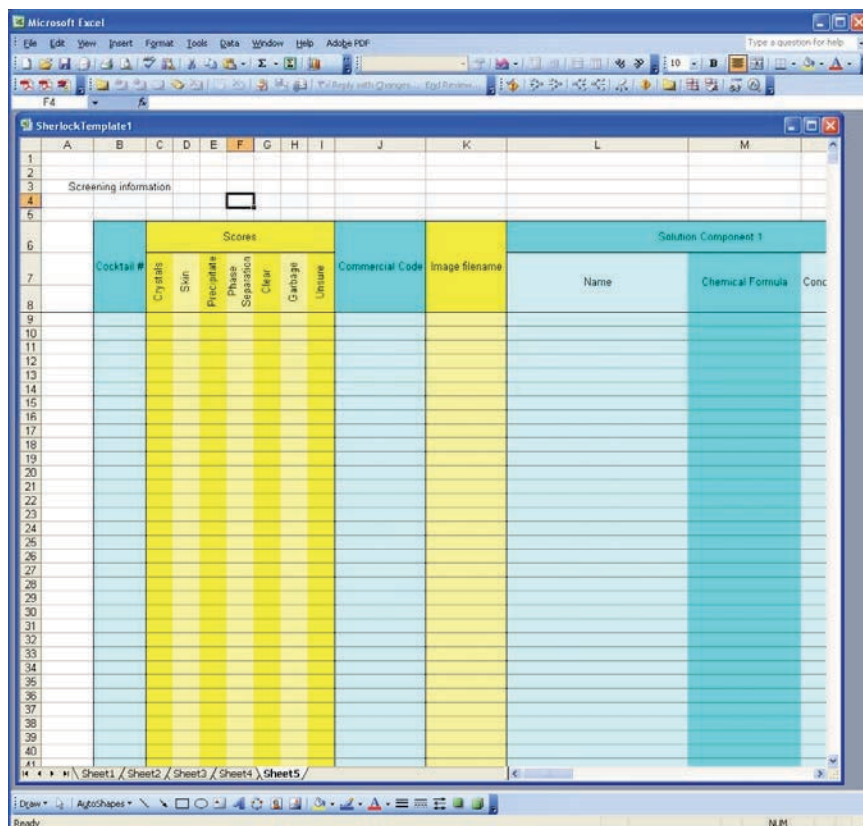
In the case of PEG, the M option represents the concentrated salts with no PEG present.



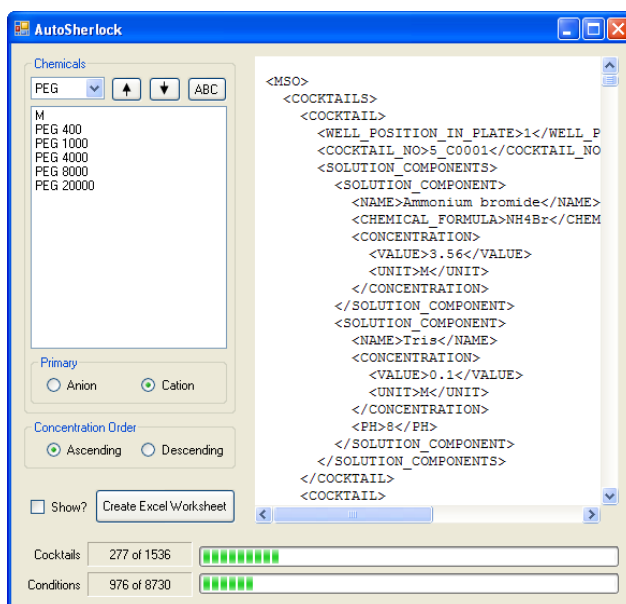
Several other options are available from this window. The default is to map chemical space with the primary list sorted according to Cation. This can be changed to Anion. Similarly, the list is sorted in order of ascending concentration. This can be changed to descending.

The final step from this point is to 'Create the Excel Worksheet'. There is an option to Show this as the worksheet is created. It is faster not to show it but we will select it here to show the workings.

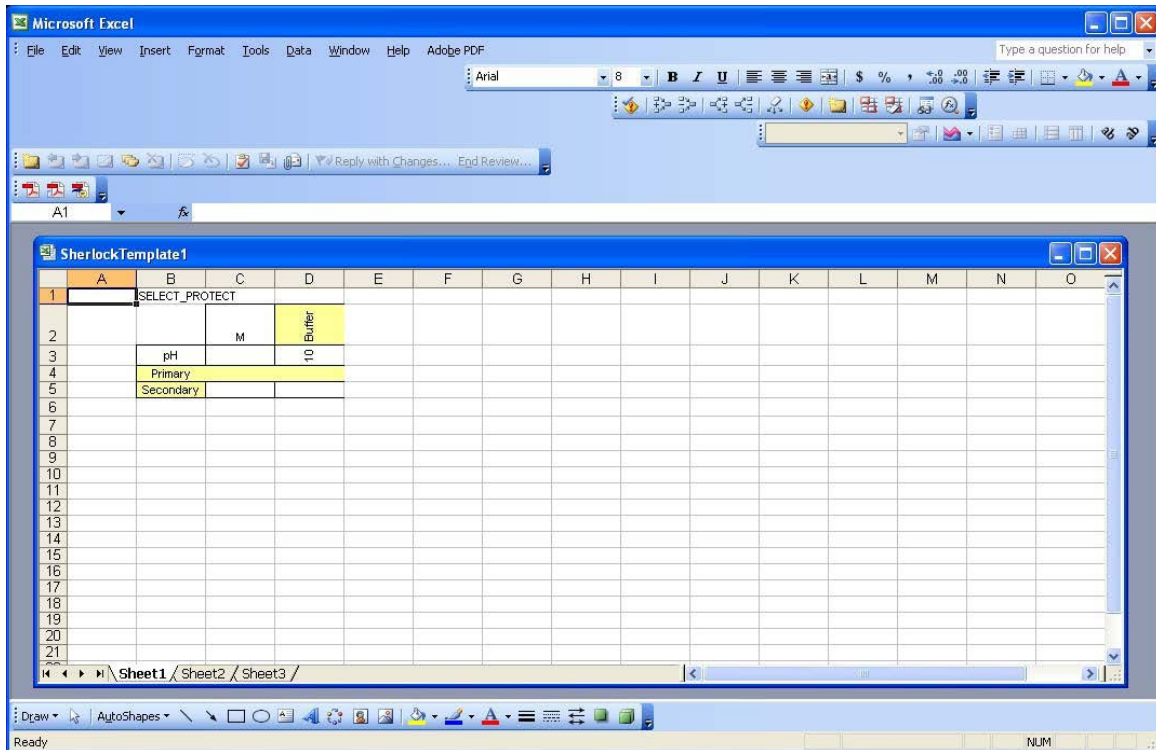
As soon as you select 'Show?' the worksheet window is visible. First time users should select 'Show' to see how the program is progressing. It will run much faster without the Excel Spreadsheet visible. The first page shown is a template for the list of chemicals and image names. Once you click 'Create the Excel Worksheet' this will be populated.



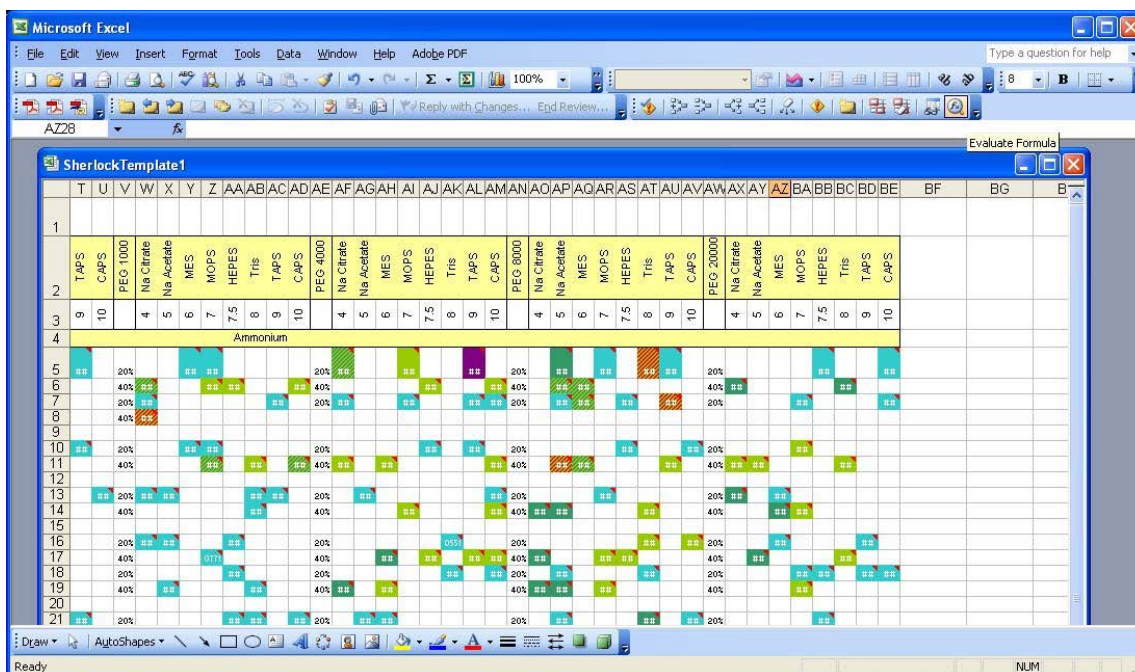
As Sherlock runs the progress bar will show it working. Initially there is no progress shown as AutoSherlock reads in the image names and creates the spreadsheet format. Once this is done the cocktails and condition counter will gradually increase.



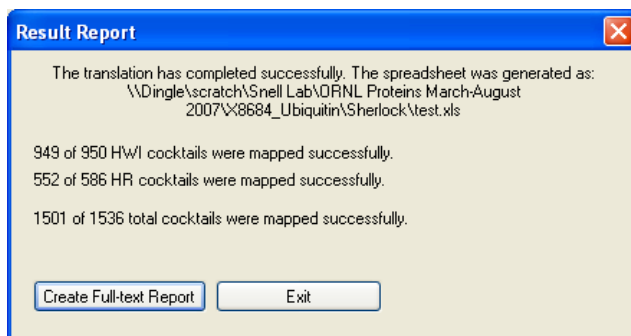
The initial spreadsheet is shown below.



The program then proceeds to populate the spreadsheet. It will alternate between three sheets writing the image names first, doing a rapid analysis of scored results then populating the main spreadsheet. **WARNING - DO NOT SELECT ANYTHING IN THE SPREADSHEET UNTIL IT HAS FINISHED.**



When AutoSherlock has finished populating the spreadsheet it will provide information on the number of cocktails identified and mapped. **NOTE - APPROXIMATELY HALF WAY THROUGH THE PROCESSING IT WILL LOOK AS IF SHERLOCK HAS PAUSED. IT HAS NOT.**



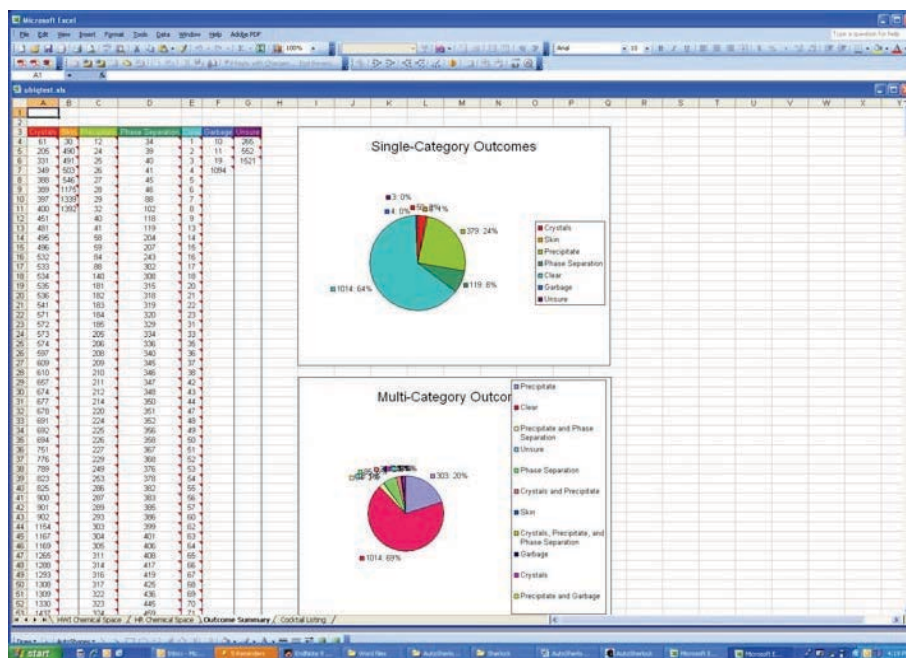
The current version only tackles the HWI cocktails and has several bugs remaining resulting in missing some of them. Both of these limitations are currently being addressed. By selecting 'Create Full-text Report' the unmapped cocktails are listed. This will ask for a filename and create an ASCII text format report on the unmapped cocktails.

7. The Sherlock spreadsheet.

7.1 The Outcome Summary

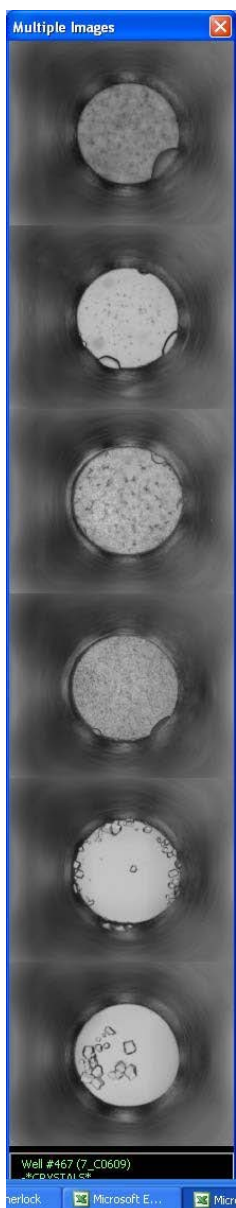
NOTE - THE SPREADSHEET USES MACROS. YOU WILL HAVE TO ACCEPT THE USE OF THESE TO ACCESS ALL THE TOOLS IN THE SPREADSHEET.

The spreadsheet contains four tabs labeled HWI Chemical Space, HR Chemical Space, Outcome Summary and Cocktail Listing. It opens with Outcome Summary first.

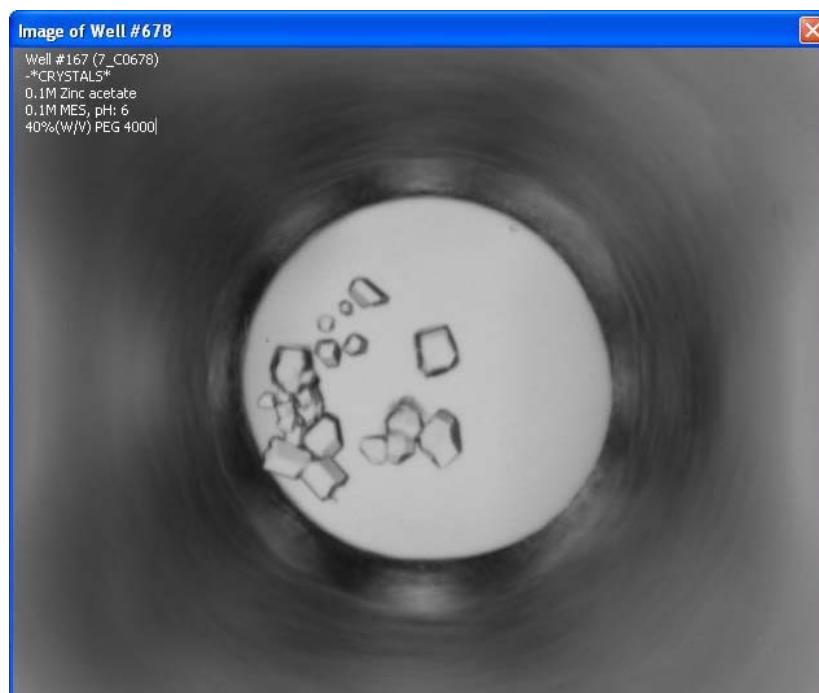


For a scored file results are color coded according to a rainbow scheme with Red representing crystals down to purple for unsure. Only the scored images are presented in the table.

Selecting these images brings up the images themselves. For example if we select 27-32 in the crystal list the following image appears;

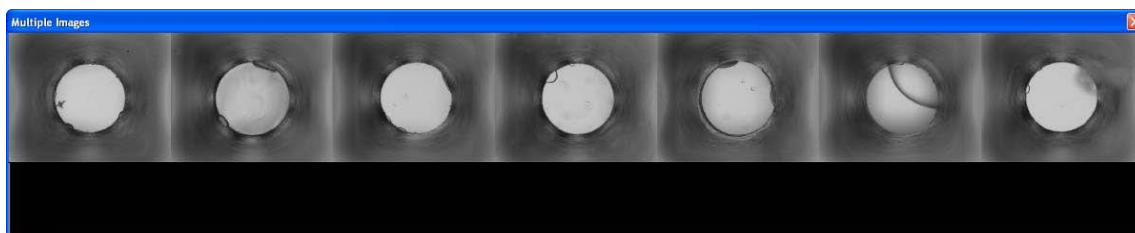


As you mouse over each image, the chemical information of that experiment is listed below. Clicking on any of the images produces a further enlargement.

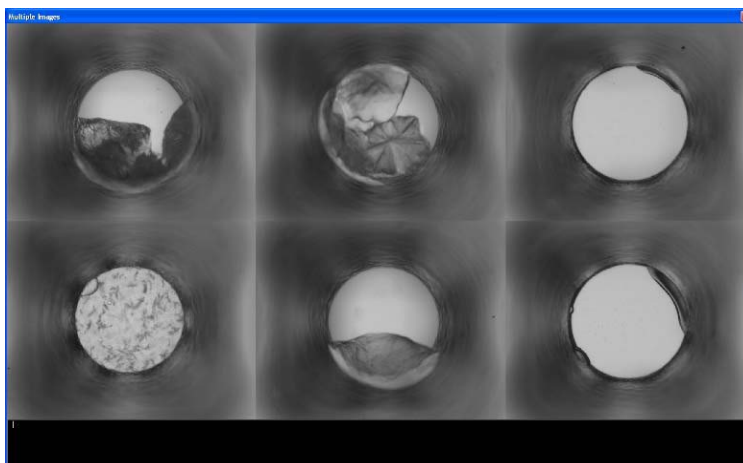


Before you can open more image, you must close all the images by clicking on the red box with the white X in the top right corner of each window that pops up.

You can select multiple images across categories, i.e. from the top row;



Multiple levels can be shown, in this case the images on the left were classified as possible crystals, the middle as skin and the ones right precipitate. This also serves as a quick check on the accuracy of scoring!



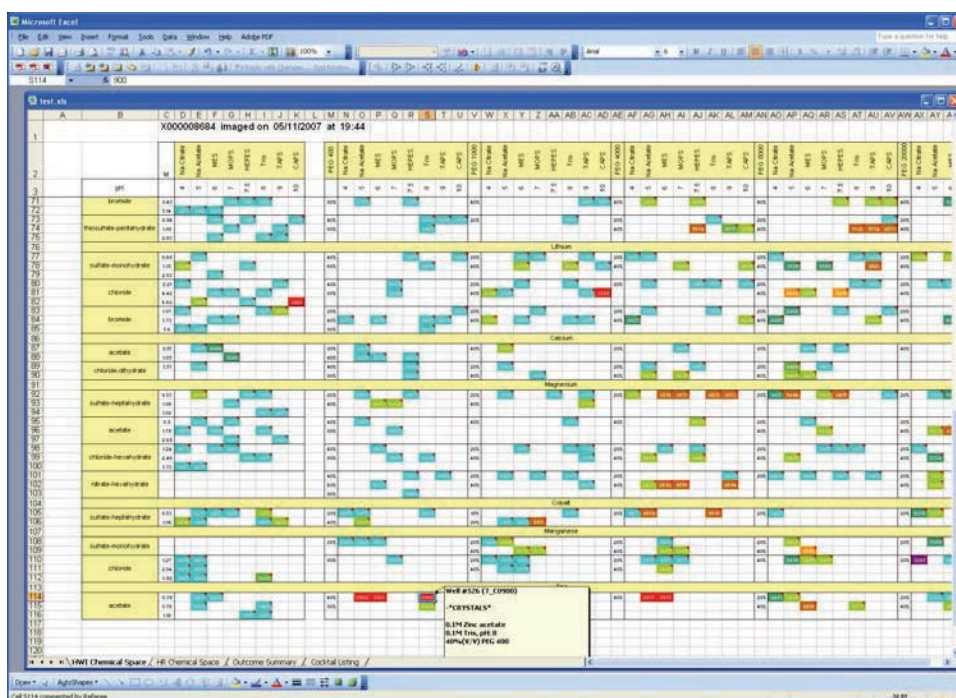
7.2 The HWI and Hampton Research Chemical Space Summary

The main usefulness of the spreadsheet for crystallization optimization is in the chemical space representation shown by clicking on the tab at the bottom of the sheet.

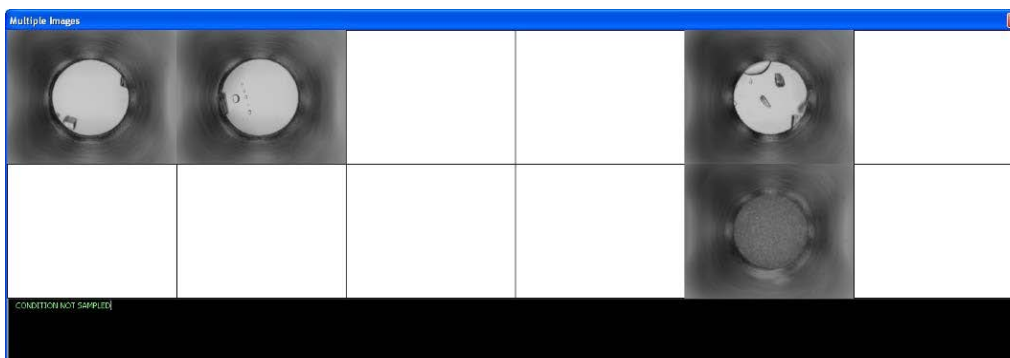
This is a large spreadsheet color coded as the first example. Multiple scores are represented with hatching of the multiple score color. To see the images you can click on any of these colored blocks. Note for un-scored results the blocks will

appear grey. White blocks represent areas of the matrix that have not been sampled in experiment. Mousing over any of the blocks will give the score and chemical conditions.

NOTE: Clicking on a cell a second time will not bring up the image again until you have deselected the cell (by selecting another) and then select it again. There is one cell that can be edited on the spreadsheet. This is cell A1 of the HWI chemical space. This contains the image directory. If the images are moved this should be changed so that Excel knows where to find the images. If images are not visible on clicking on a cell it is likely that A1 contains the incorrect image directory.



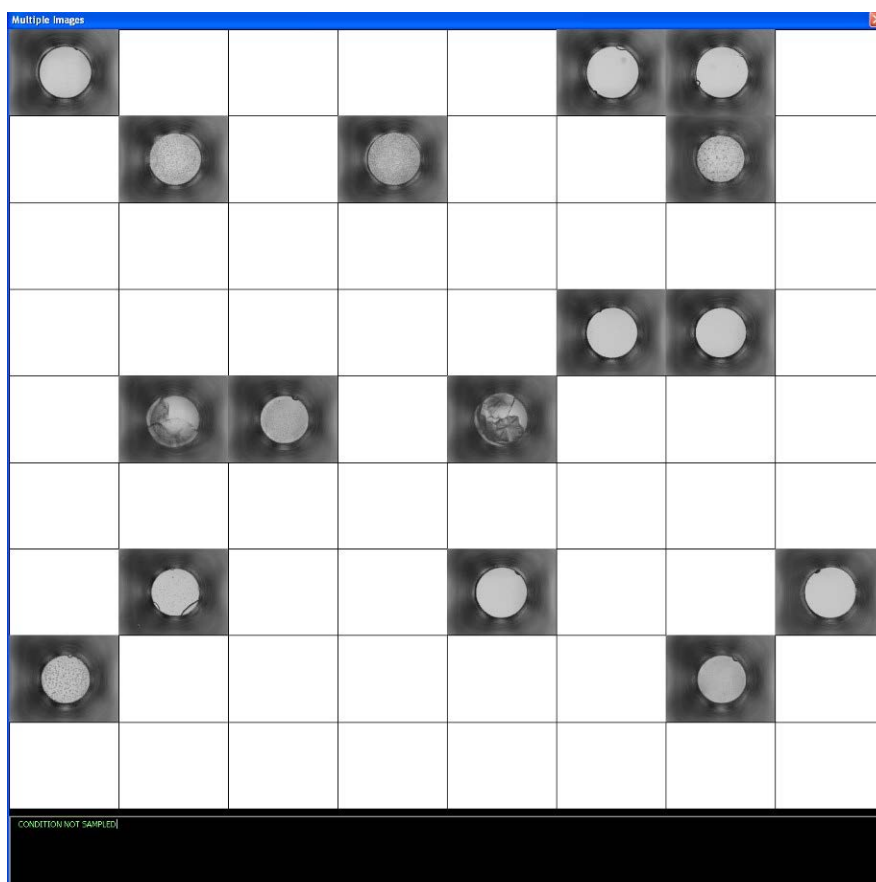
Selecting a block will show all those images in the block along with white space where there are no experiments.



This can identify directions for optimization. Mousing over any of the pictures presents chemical information on that picture. Clicking on any picture provides an enlargement of that picture e.g.



It is possible to select many images with the number only limited by the computers capabilities e.g.



The Hampton Research Cocktails in the HWI 1536 screen are shown on the second tab of the spreadsheet, 'HR Chemical Space'.

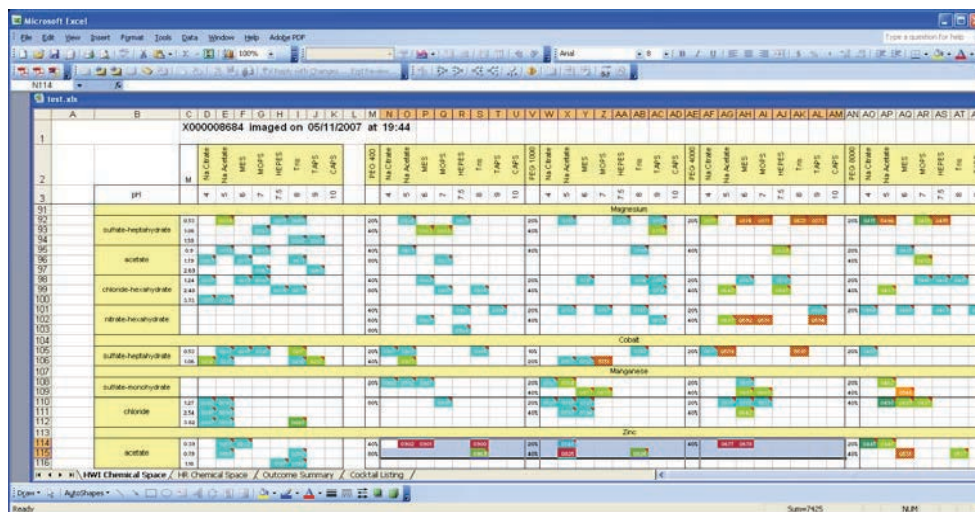
8. How to Use Sherlock

Sherlock is intended to be used for several purposes:

1. A quick look at how the results relate to each other.
2. To see what conditions have not been sampled.
3. To identify areas where crystals may not have been scored correctly.
4. To provide direction for optimization.

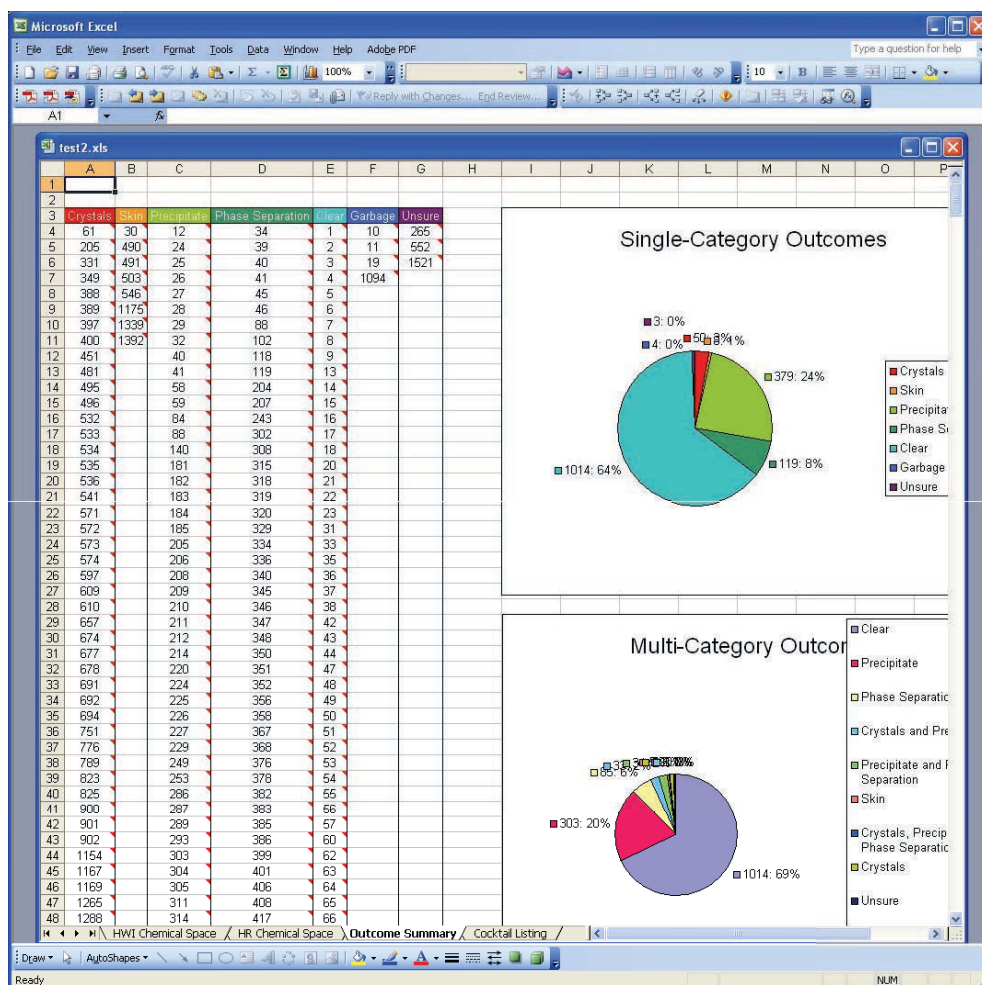
8.1 How results relate to each other.

The results are formatted in chemical space. The pH, PEG concentration if any, and salt information are given for all the HWI incomplete factorial screen. For scored images you can quickly identify what components are producing crystals. In the example case, looking at the 'HWI Chemical Space' tab, zinc acetate and low molecular weight PEGs seem to be a clear area of crystal leads. Calcium and magnesium acetate do not seem to be giving similar results so we can hypothesize that zinc is important in this case.

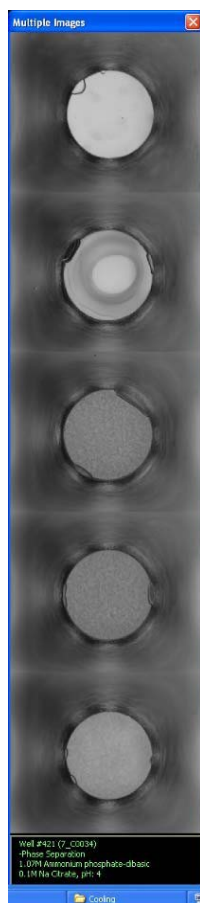


Note: Over the years, conditions that produce salt crystals have been slowly eliminated from the HWI incomplete factorial but it is still possible that salt crystals could form.

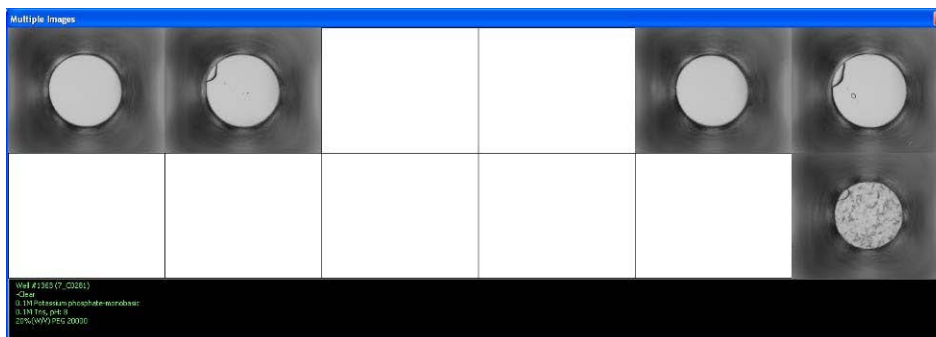
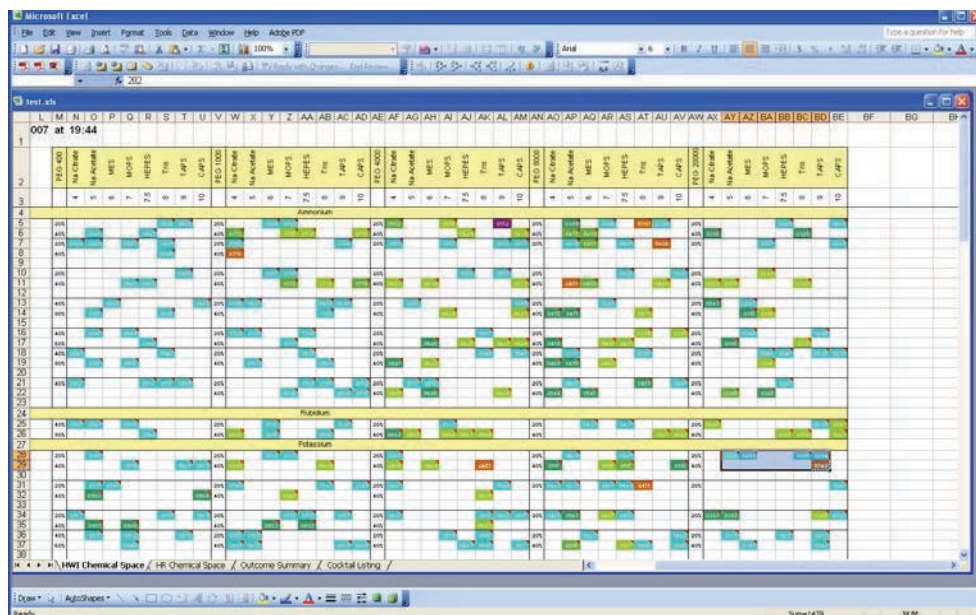
If we switch to the 'HR Chemical Space' tab we can confirm that zinc acetate is giving a hit in the Hampton Index index screen and also in the Hampton Index Crystal Screen HT. Interestingly, in the case of the HT screen one of the hits is from zinc acetate but the other is from sodium acetate which has not been sampled as part of the HWI incomplete factorial shown on the first tab.



For example, if we select the first 5 phase separation examples we might want to reclassify numbers 40 and 41 as precipitate. Checking the precipitate list we see that they are already classified as this as well as phase separation.



The other way to check scores is to use the Chemical Space tabs to locate examples that appear out of place. For example, if we look at Potassium phosphate monobasic, PEG 20K, we see something at pH 9 but nothing elsewhere.



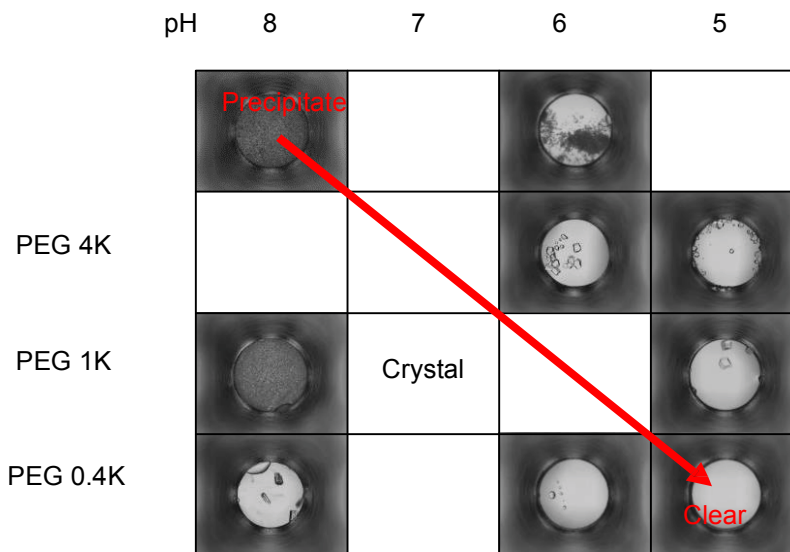
Have we scored this correctly? We can gain some confidence in the score by looking at the results in other PEGs, i.e. 8K and 4K. There is indeed precipitation seen and possibly hints of crystals. Interestingly this particular case shows phase separation at pH 10, PEG 8K. Phase separation may be an indication of potential crystallization conditions.

8.4 Optimization

The strength of Sherlock comes in the optimization stage. Rational optimization requires two elements, an idea of the variables that drive optimization and the ranges of those variables. In this particular case, we have identified PEG 400-8000K, 20%-40%, pH 5-8 and zinc acetate from the screen. We also know that outside of this we have not sampled pH 4 or lower and higher PEG concentrations (except in a single case). This provides our arena for optimization.

Where we do not have a clear case of crystals we can choose conditions that consist of clear and precipitate separated by un-sampled space. If crystallization is to occur it will occur on a vector between the two.

We can take the optimization one stage further by extracting data from Sherlock. For example, we can construct the following vector map for the zinc acetate conditions. It is clear that pH 7 is a promising but un-sampled area to look at, especially at PEG 4-1K and points in between. Interestingly, the zinc acetate concentration is not a variable that has been sampled.



8.5 Other features

At the bottom of the 'Outcome Summary' list are statistics that describe the breakdown of scores.

	A	B	C	D	E	F	G	H	I	J
1012					1530					
1013					1531					
1014					1532					
1015					1533					
1016					1534					
1017					1535					
1018	50	8	379	119	1014	4	3			
1019										
1020	Precipitate							303		
1021	Clear							1014		
1022	Precipitate and Phase Separation							34		
1023	Unsure							3		
1024	Phase Separation							85		
1025	Crystals and Precipitate							33		
1026	Skin							5		
1027	Crystals, Precipitate, and Phase Separation							5		
1028	Garbage							3		
1029	Crystals							12		
1030	Precipitate and Garbage							1		
1031	Skin, Precipitate, and Phase Separation							3		
1032										
1033										
1034										
1035										
1036										
1037										
1038										

As a guide, we have observed more crystallization success in cases where the ratio of precipitant to clear is greater than 1. Less may indicate that greater sample concentration is needed or that precipitant concentration should be increased.

9. *Known Bugs*

There are plenty but we'd like to hear about from them from you - please email rnagel@hwi.buffalo.edu or esnell@hwi.buffalo.edu.

10. Unknown Bugs

See Known Bugs :)

11. Acknowledgements and references.

Implementation and coding Ray Nagel (rnagel@hwi.buffalo.edu). Layout and concept, Eddie Snell (esnell@hwi.buffalo.edu).

AutoSherlock references:

Nagel RM, Luft, JR, Snell, EH. (2008). AutoSherlock: a program for effective crystallization data analysis *J Appl Crystallogr*, **41**, 1173-1176.

Snell EH, Nagel RM, Wojtaszyk A, O'Neill H, Wolfley JR, Luft JR. (2008) The application and use of chemical space mapping to interpret crystallization screening results *Acta Crystallogr D Biol Crystallogr*, **64**, 1240-1249.

The HWI crystallization screening laboratory reference:

Luft, J. R., Collins, R. J., Fehrman, N. A., Lauricella, A. M., Veatch, C. K. & DeTitta, G. T. (2003). A deliberate approach to screening for initial crystallization conditions of biological macromolecules. *J. Struct. Biol.* **142**, 170-179.