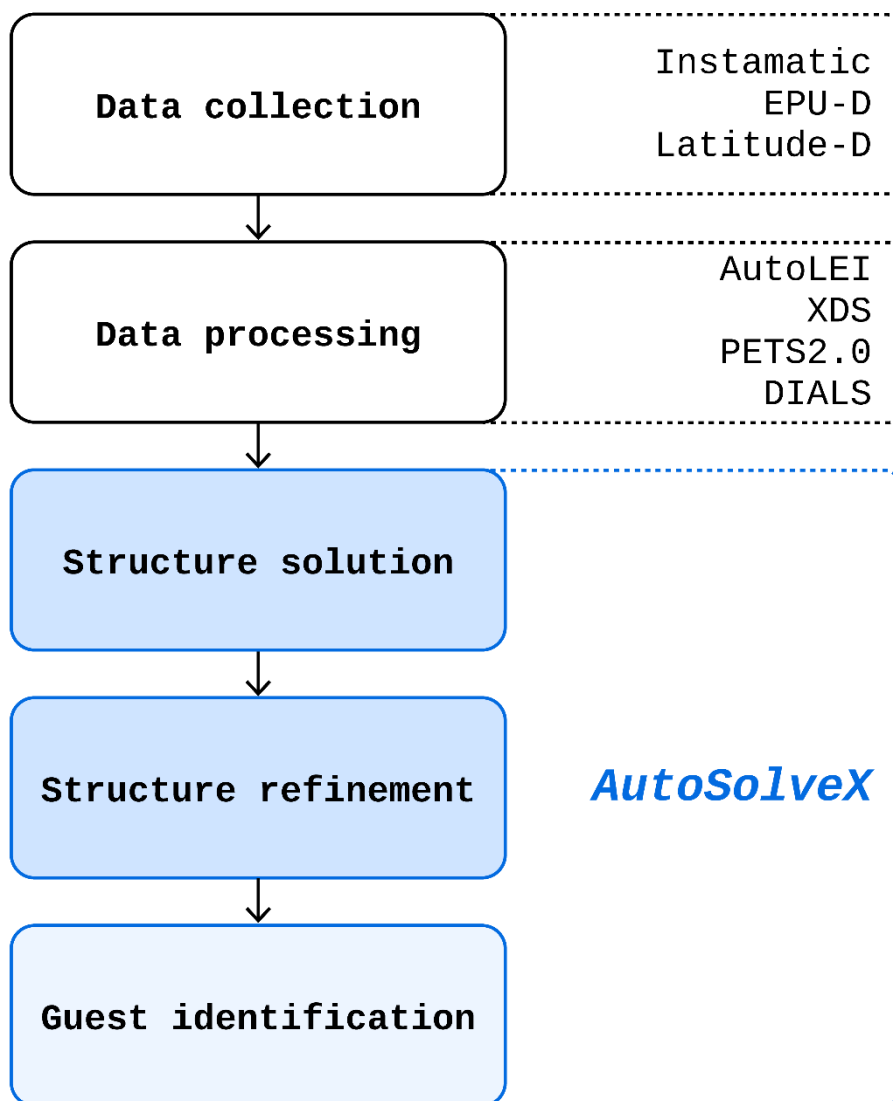


Tutorial for AutoSolveX

AutoSolveX: a Python-based pipeline for automated structure solution, refinement, and guest identification from diffraction data

Sofiia Butonova, Yinlin Chen, Xiaodong Zou



Examples of structure determination and guest-molecule identification in porous materials using AutoSolveX, together with details of the underlying algorithms, are provided in the following publication:

<https://chemrxiv.org/doi/full/10.26434/chemrxiv.10002125/v1>

Contents

Installation	3
Troubleshooting.....	4
Functionality of the Tabs.....	5
Solution and Refinement.....	5
1. Initial steps	5
1.1 Select the project folder	5
1.2 Enter the space group and composition	6
2. General workflow.....	6
2.1 Run structure solution	6
2.2 Update reflection files.....	6
2.3 Prepare .ins files for refinement.....	6
2.4 Run refinement.....	7
3. Model-based workflow	7
3.1 Switch to model mode	7
3.2 Select a model .res file	7
3.3 Update the model .ins	7
3.4 Update .hkl files	8
3.5 Refine the model.....	8
4. Inspect output files.....	8
5. Collect statistics	8
Guest Identification	9
1. Select guest molecules	9
2. Run guest identification	10
3. Adjust guest-identification settings.....	10
4. Visualize selected Q-peaks	10
5. Assign Q-peaks and refine the guest model	10
6. Show refinement and occupancy results	11
7. Save the current table	11
Opening structures in Olex2	12
Cleaning generated files.....	12

Installation

Common procedure on Windows with Python ≥ 3.8 (single Python install) to install your local autosolvex-0.1.0.251216.tar.gz and make autosolvex runnable from cmd.

Installing AutoSolveX on Windows

- 1) Verify you are using Python ≥ 3.8

`python --version` (must show Python ≥ 3.8 . If not, you need to install Python first)

- 2) Upgrade pip/tools (recommended for sdist)

`python -m pip install --upgrade pip setuptools wheel`

- 3) Install your tar.gz

Go to the folder with the file and install the latest version:

`python -m pip install autosolvex-0.1.0.251216.tar.gz`

- 4) Ensure the autosolvex command is runnable from the command prompt (cmd):

`autosolvex` (If cmd says that autosolvex is not recognized, the Python Scripts folder is probably not added to PATH. You can find details in the Troubleshooting part)

- 5) You can also check where the command is located:

`where autosolvex`

- 6) Ensure SHELX is on PATH:

`where shelxt` (If shelxt is found, cmd will show the path to shelxt.exe.)

If not found, add the folder containing shelxt.exe to PATH the same way.

Alternative installation using pip from PyPI

AutoSolveX is available as a pip package, it can be installed directly with:

`python -m pip install autosolvex`

After installation, AutoSolveX can be started from any directory by running:

`autosolvex`

To uninstall AutoSolveX, run:

`pip uninstall autosolvex`

Troubleshooting

1. autosolvex is not recognized

If cmd says:

```
'autosolvex' is not recognized as an internal or external command
```

then the Python Scripts folder is probably missing from PATH.

First, find your Python installation path:

```
python -c "import sys; print(sys.executable)"
```

For example, if the output is:

```
C:\Users\<USER>\AppData\Local\Programs\Python\Python38\python.exe
```

then the corresponding Scripts folder is usually:

```
C:\Users\<USER>\AppData\Local\Programs\Python\Python38\Scripts
```

If the output is:

```
C:\Program Files\Python38\python.exe
```

Solution: Add the Python Scripts folder to PATH on Windows

- Press Win and search for Environment Variables.
- Open Edit the system environment variables.
- Click Environment Variables....
- Under User variables, select Path and click Edit.
- Click New.
- Paste the Python Scripts folder path.
- Click OK until all windows are closed.
- Open a new cmd.
- Test again:

```
where autosolvex
```

```
autosolvex
```

Functionality of the Tabs

After launching AutoSolveX, the main window opens with two tabs:

- Solution and Refinement (Figure 1)
- Guests identification (Figure 2)

The **Solution and Refinement tab** is used to prepare SHELX input files, run structure solution, update reflection files, refine structures, and collect statistics.

The **Guests identification tab** is used to identify possible guest molecules from Q-peaks, visualize selected Q-peaks, assign Q-peaks to guest atoms, and refine guest-containing models.

A log window is shown at the bottom of the GUI. Important messages, warnings, progress information, and SHELX output are displayed there.

Solution and Refinement

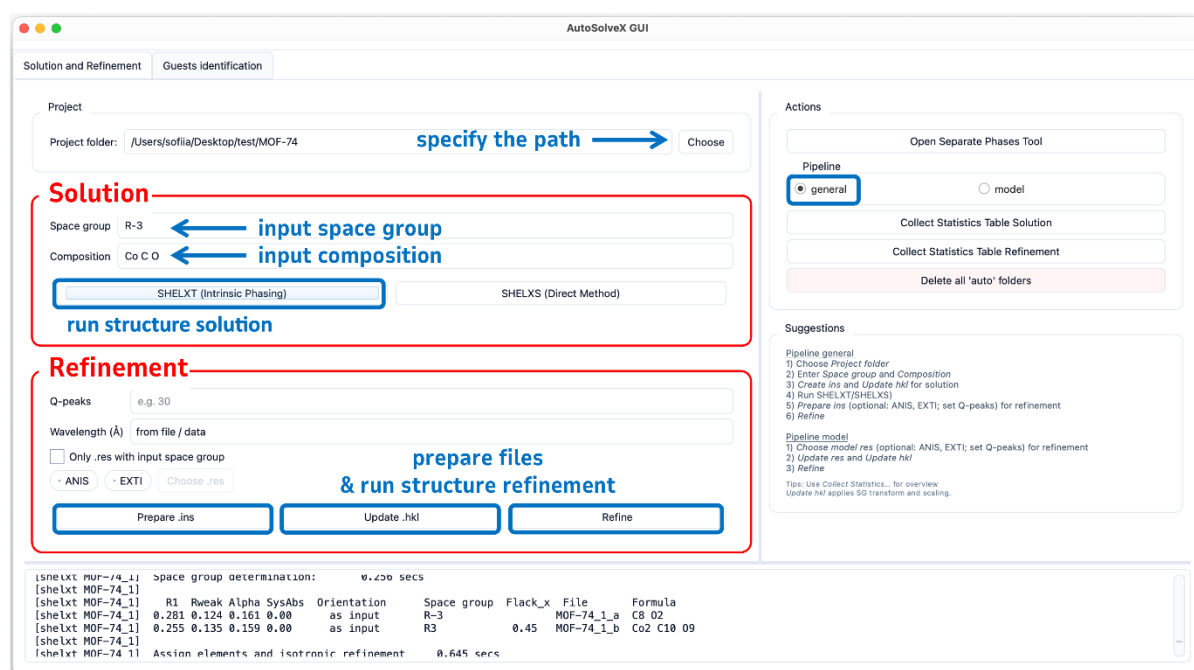


Figure 1 Solution and Refinement Tab.

1. Initial steps

1.1 Select the project folder

Click **Choose** in the **Project** section and select the parent project folder.

This folder should contain the structure folders that will be processed by AutoSolveX. After the folder is selected, its path is shown in the **Project folder** field.

AutoSolveX uses this folder as the working directory for solution, refinement, statistics collection, and guest identification.

1.2 Enter the space group and composition

In the **Solve (.ins)** section, enter:

- **Space group**
Example:
R-3
- **Composition**
Example:
Co C O

The composition is used to prepare the SHELX input files and to compare the expected composition with the composition extracted from SHELXT results. The GUI contains a mismatch label that can show possible element substitutions when the parsed SHELXT composition does not match the user input.

2. General workflow

2.1 Run structure solution

Two structure-solution options are available:

- **SHELXT (Intrinsic Phasing)**
- **SHELXS (Direct Method)**

Click **SHELXT (Intrinsic Phasing)** for the standard intrinsic phasing workflow.

Click **SHELXS (Direct Method)** to use the direct-method workflow.

When one of these buttons is clicked, AutoSolveX creates or updates SHELX.INS files from the available P4P files and then prepares the corresponding HKL files before running SHELX. The code searches for P4P files and uses the input space group and composition to write SHELX input files.

2.2 Update reflection files

Click **Update HKL** to update the reflection files.

This step applies the required space-group transformation when needed. In solution mode, AutoSolveX creates or refreshes HKL files for the corresponding .ins files.

Use this step before refinement if the reflection files need to be regenerated or transformed.

2.3 Prepare .ins files for refinement

In the **Prepare / Refine** section, set the refinement options before clicking **Prepare .ins**.

Available options include:

- **Q-peaks**
Number of Q-peaks to keep in the SHELX file.

- **Wavelength (Å)**

Available options include:

- from file/data
- 0.0197 Å for 300 kV 3D ED
- 0.0251 Å for 200 kV 3D ED
- 0.71073 Å for Mo K α SCXRD
- 1.54 Å for Cu K α SCXRD

- **Only .res with input space group**

If this option is checked, only .res files whose space group matches the input space group are prepared and refined.

- **ANIS**

Adds the ANIS instruction for anisotropic refinement of atoms.

- **EXTI**

Adds the EXTI instruction for refinement of the extinction parameter.

Click **Prepare INS** to generate updated .ins files in the corresponding auto folders. The preparation step uses the selected composition, wavelength, Q-peak number, and optional ANIS/EXTI settings.

2.4 Run refinement

Click **Refine** to run SHELXL refinement.

AutoSolveX runs SHELXL on the prepared files, extracts the recommended WGHT, copies the updated RES file back to .ins, and continues refinement until convergence. The progress bar shows the number of structures being refined and allows the run to be canceled.

3. Model-based workflow

AutoSolveX also supports refinement using an existing model.

3.1 Switch to model mode

In the **Pipeline** section, select: **model**

3.2 Select a model .res file

Click **Choose .res** and select the SHELX .res file that will be used as the model.

AutoSolveX reads information from this file, including Q-peaks, wavelength, and the presence of ANIS or EXTI instructions. These values are then reflected in the GUI settings.

3.3 Update the model .ins

Adjust the settings if needed:

- Q-peaks
- wavelength
- ANIS

- **EXTI**

Then click **Update .ins**.

This creates or updates an **INS** file from the selected model **RES** file and stages the model files into the corresponding auto folders.

3.4 Update .hkl files

Click **Update HKL** to stage and transform the reflection files for model refinement.

3.5 Refine the model

Click **Refine**.

AutoSolveX runs SHELXL on the staged model jobs and performs the same convergence workflow as in the general refinement pipeline.

4. Inspect output files

After refinement, the refined files are written to the corresponding auto folders. The main files to inspect are the final **RES** file and the **LST** refinement log.

To inspect the results in the GUI, click **Collect Statistics Table Refinement**, select a structure from the table, and click **Open Selected in Olex2**. Then check the refinement statistics, residual density, displacement parameters, geometry, and possible warnings in the **LST** file.

5. Collect statistics

Two statistics tools are available in the **Actions** panel:

- **Collect Statistics Table Solution**
- **Collect Statistics Table Refinement**

Use **Collect Statistics Table Solution** to collect solution statistics from **LXT** files.

Use **Collect Statistics Table Refinement** to collect refinement statistics from valid **CIF/FCF** outputs.

The statistics tables can be saved as **.xlsx** or **.csv** files, and selected structures can be opened in **Olex2** directly from the table.

Guest Identification

The **Guests identification** tab (Figure 2) is used to compare Q-peak clusters against possible guest molecules.

Guest identification

Q-peaks: 50 ← specify number of Q-peaks

Find molecules...

- ✓ (R)-2,3-dimethoxy-5,10,12,12a-tetrahydropyrazolo[3,4'-4,5]pyrido[2,1-a]isoquinolin-8(6H)-one, R-32 (COC1=C(O)C=C(CCN2(C@H)3CC4=NNC=C4C2=O)C3=C1)
- ✓ (S)-2,3-dimethoxy-5,10,12,12a-tetrahydropyrazolo[3,4'-4,5]pyrido[2,1-a]isoquinolin-8(6H)-one, S-32 (COC1=CC2=C(CCN3(C@H)2CC4=NNC=C4C3=O)C=C1OC)
- ✓ 1H-1,2,3-triazole, 18 (N1N=NC=C1)
- ✓ 1H-1,2,4-triazole, 19 (N1N=CN=C1)
- ✓ 1H-indazole-5-boronic acid pinacol ester, 30 (C1C1(OB(C2=CC3=C(NN=C3)C=C2)OC1(C)C(C)C)
- ✓ 1H-indazole, 28 (C12=C(NN=C2)C=C(C=C1)
- ✓ 1H-pyrazole, 17 (N1N=CC=C1)
- ✓ 4-bromo-3-methyl-1H-pyrazole, 21 (CC1=NNC=C1Br)
- ✓ 5-methyl-2H-tetrazole, 20 (CC1=NN=NN1)
- ✓ cyclohexanesulfonic acid, 29 (O=S(=O)(CCCC1)O)
- ✓ dimethyl phosphate, 26 (O=P(OC)(O-)(OC)
- ✓ dimethylphosphinic acid, 26 (O=P(C)(C)O)
- ✓ Estrodiol, 31 (C18H26O2)C18H26O2=NC(C4=CC=C(C(C=C5C6=C4)C=NC6C3=NC=C6(C=C1)

Enter SMILES: e.g. C1=CC=CC=C1 Add

run guest identification

Actions

- Settings
- Open Selected in Olex2
- Visualize Selected in 3D
- Assign in Selected Q to C and refine
- Show Q-peaks Table
- Show Occupancy Refinement Table
- Save table (xlsx)

Guests table

file	rank	guest	s	RMSD (Å)	Nq	Nmol	Q-peaks labels used
delta-MOF-520-17	1	17	0.99	0.07	5	5	Q1_7,Q2_0,Q3_7,Q4_0,Q5_7
delta-MOF-520-17	2	18	0.98	0.07	5	5	Q1_7,Q2_0,Q3_7,Q4_0,Q5_7
delta-MOF-520-17	3	19	0.98	0.08	5	5	Q1_7,Q2_0,Q3_7,Q4_0,Q5_7
delta-MOF-520-17	4	20	0.83	0.26	6	6	Q1_4,Q2_1,Q3_4,Q4_1,Q5_4,Q19_0
delta-MOF-520-17	5	28	0.4	1.62	11	9	Q1_2,Q1_5,Q2_3,Q2_6,Q3_2,Q3_5,Q4_3,Q4_6,Q5_2,Q5_5,Q19_2
delta-MOF-520-17	6	21	0.4	1.83	9	7	Q1_1,Q1_6,Q2_2,Q2_4,Q3_1,Q3_6,Q4_2,Q4_4,Q5_1
delta-MOF-520-18	1	18	1.01	0.14	5	5	Q1_4,Q2_4,Q3_4,Q4_4,Q5_1

check the table

=== delta-MOF-520-17 : Top-6 candidates ===

1. 17 rmsd=0.074 N_q/N_mol=5/5 used=Q1_7,Q2_0,Q3_7,Q4_0,Q5_7
2. 18 rmsd=0.074 N_q/N_mol=5/5 used=Q1_7,Q2_0,Q3_7,Q4_0,Q5_7
3. 19 rmsd=0.082 N_q/N_mol=5/5 used=Q1_7,Q2_0,Q3_7,Q4_0,Q5_7
4. 20 rmsd=0.256 N_q/N_mol=6/6 used=Q1_4,Q2_1,Q3_4,Q4_1,Q5_4,Q19_0
5. 28 rmsd=1.619 N_q/N_mol=11/9 used=Q1_2,Q1_5,Q2_3,Q2_6,Q3_2,Q3_5,Q4_3,Q4_6,Q5_2,Q5_5,Q19_2
6. 21 rmsd=1.834 N_q/N_mol=9/7 used=Q1_1,Q1_6,Q2_2,Q2_4,Q3_1,Q3_6,Q4_2,Q4_4,Q5_1

Figure 2 Guest Identification Tab.

1. Select guest molecules

The guest list is loaded from the internal guest molecule database. You can search for molecules using the filter box.

Check the molecules that should be included in the search.

If no molecule is selected, AutoSolveX uses all available guests. Add a custom guest molecule

To add a custom molecule:

1. Enter a SMILES string in the **Enter SMILES** field.
2. Click **Add**.

The molecule is added to the guest list and checked by default.

Example:

C1=CC=CC=C1

2. Run guest identification

Optionally enter the number of Q-peaks to use. Otherwise, the program will use all Q-peaks from RES files.

Then click: **Run guest identification**

AutoSolveX runs guest identification on the selected project folder and displays the top candidate guest molecules in the table.

3. Adjust guest-identification settings

Click **Settings** to adjust the guest-identification parameters:

- **Peak height cutoff (τ):** Keep only Q-peaks with intensity $I_{Qn} \geq \tau I_{Q1}$ in the res file (higher τ keeps fewer peaks).
- **Q-Q distance cutoff (Å):** Maximum Cartesian distance for adding an edge between two Q-peaks in the Q-Q graph and considering this as a Q-Q bond; this directly affects topology and ring detection.
- **Max Q-peaks per file for screening (0 = all):** Use only the first N Q-peaks from each RES before intensity and distance-based filtering.
- **Max extra Q-peaks above $N_{m \text{ non-H}}$:** Maximum allowed extra Q-peaks above $N_{m \text{ non-H}}$ (0 = enforce $N_q \text{ used} = N_{m \text{ non-H}}$), depending on the target molecule size.
- **Q-peak clustering cell setting: single unit cell and supercell $2 \times 2 \times 2$.** In the **single unit cell** mode, only Q-peaks located within one unit cell are considered. In the **$2 \times 2 \times 2$ supercell mode**, Q-peaks are expanded into a $2 \times 2 \times 2$ supercell before clustering and guest-molecule matching.
- **Rigid or non-rigid** treatment of 5- and 6-membered rings. In the **non-rigid** mode, a 5-membered ring is allowed to have one extra internal Q-Q connection between two atoms that should not form a chemical bond. In the **rigid** mode, extra internal Q-Q connections are not allowed.
- **Framework-Q-peak distance cutoff factor (related to framework atom type):** For each observed element in the datasets, define a minimum allowed Q-framework distance as (van der Waals radius) $\times$ (Framework - Q-peak distance cutoff factor); any Q-peak closer than this threshold to its nearest framework atom is filtered.

After clicking **OK**, the new settings are used for the next Q-peak analysis.

4. Visualize selected Q-peaks

Select a row in the guest table and click: **Visualize Selected in 3D**

AutoSolveX opens a 3D visualization of the selected structure and highlights the Q-peaks used for the selected guest candidate. If Qt WebEngine is available, the plot is shown inside the GUI. Otherwise, AutoSolveX falls back to a static image or opens the plot in the browser.

5. Assign Q-peaks and refine the guest model

Select one or more rows in the guest-identification table.

Then click: **Assign in Selected Q to C and refine**

AutoSolveX reads the Q-peak labels from the selected row, assigns them to guest carbon atoms, creates new guest .ins and .hkl files, and runs SHELXL refinement only for the newly created guest jobs. The generated files are named using the guest name and the original structure name, for example: *<guest>@<structure>.ins*

If the corresponding guest job already exists, AutoSolveX does not overwrite it automatically and reports that no new refinement job was created.

6. Show refinement and occupancy results

After guest refinement, click:

Show Occupancy Refinement Table

This displays a table containing guest-refinement results. When available, the table includes:

- identifier
- guest
- guest occupancy
- space group
- resolution
- completeness
- R1
- wR
- maximum shift
- path to the RES file

7. Save the current table

Click: **Save table**

This saves the currently displayed guest table as an xlsx file. The exported file preserves the visible table content, headers, current sorting, and approximate GUI styling.

Opening structures in Olex2

Several tables include an **Open Selected in Olex2** button.

To use it:

1. Select a row in the table.
2. Click **Open Selected in Olex2**.

AutoSolveX opens the corresponding RES or INS file in Olex2.

Cleaning generated files

The **Solution and Refinement** tab contains the button: **Delete all 'auto' folders**

Use this only when you want to remove all automatically generated auto folders inside the selected project folder. **This operation is irreversible.**