Graz University of Technology
Institute of Solid State Physics

GIDInd Manual 2.0

This file contains the manual and two tutorials for the indexing software GIDInd.

Version 2.0 - GIDInd_2021_02_25

`https://www.if.tugraz.at/amd/GIDInd/` and

`https://github.com/m-kainz/GIDInd`.

Manual written and edited by Manuel Kainz

Software developed by Manuel Kainz and Josef Simbrunner

March 2, 2021

Abstract

Grazing-incidence X-ray diffraction (GIXD) is a widely used technique for the crystallographic characterization of thin films. The identification of a specific phase or the discovery of an unknown polymorph requires indexing of the associated diffraction pattern. Indexing describes the assignment of Laue indices to the individual diffraction peaks with simultaneous determination of the lattice constants. However, despite the importance of this procedure, only few approaches have been developed so far. Recently, a mathematical framework for indexing of this specific diffraction pattern has been developed. This work introduces a successfully implemented indexing routine. The algorithm is based on the assumption of a triclinic unit cell with six lattice constants. Two approaches are chosen: i) using only diffraction peaks of the GIXD pattern and ii) combining the GIXD pattern with a specular diffraction peak. In the first approach the six unknown lattice parameters have to be determined by a single fitting procedure, while in the second approach two successive fitting procedures are used with three unknown parameters each. The latter case is elaborated in detail throughout this thesis. The computational toolkit is compiled in the form of a MATLAB application and presented within a user-friendly graphical user interface. The program is tested by application to four independent examples of organic thin films.

Contents

Chapter 1

Instruction Manual and Tutorial

With the following script, the graphical user interface (GUI) of the indexing routine GIDInd is introduced. Please note that the current version of the manual might not match the current version of the GIDInd version used. It is tried to maintain the software and the manual up-to-date, but without any warranty. The last updated version of the manual is based on *GIDInd_2021_02_25*.

The main window consists of five different sub-windows, hereinafter referred to as *panels*. Figure 1.1 shows a screenshot of the main window with the five panels marked in different colours.

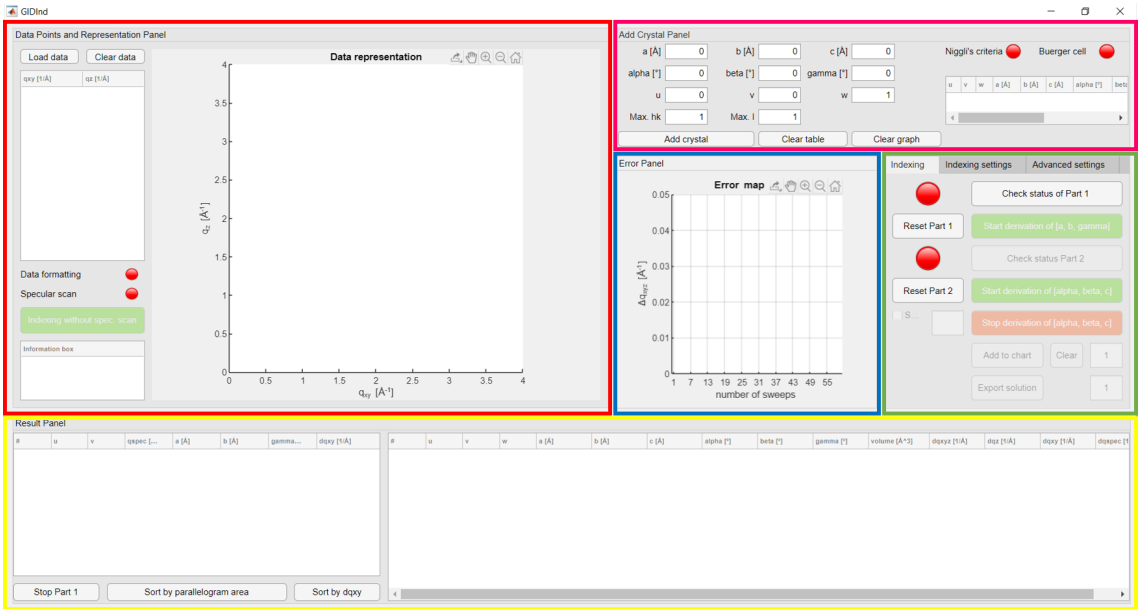


Figure 1.1: Screenshot of the indexing routine's main window right after start and prior to any input.

Below, the possible settings and features of every panel are explained individually. The

panels are indicated and described in the following order:
(i) *Add Crystal Panel* in magenta, (ii) *Data Points and Representation Panel* in red, (iii) *Indexing Panel* in green, (iv) *Error Panel* in blue and (v) the *Result Panel* indicated in yellow. For demonstration purposes, an organic thin film sample with the name "Cu INA"-MOF is used. The peak positions are derived and provided by *Lukas Legenstein* using the software tool GIDVis [1].

The program is provided in the form of an executable file (.exe) and as MATLAB application (.mlapp-file), written and tested in MATLAB version R2019b, Update 5. MATLAB Runtime or a valid licence is required. One important prerequisite to operate GIDInd source code is the MATLAB *Parallel Computing Toolbox*™. As previously mentioned, the toolbox helps to reduce computation time through dividing program blocks in smaller units. The program can be used as standalone executable file and as .mlapp file together with the functions provided in the zip-compiled folder. Beside the toolbox for parallel computing, no further installation is required for the app. The executable can be used right away, if the MATLAB Runtime is installed. To access the program, please visit `https://www.if.tugraz.at/amd/GIDInd/` and `https://github.com/m-kainz/GIDInd`.

1.1 Add Crystal Panel

The *Add Crystal Panel* can be used to simulate diffraction peaks and to check cell parameters for the reduced cell and Niggli criteria. A screenshot is shown in Figure 1.2a. The panel serves two purposes. The first is to check whether or not a cell fulfills Niggli's criteria and the criteria for a reduced cell according to Buerger. The second purpose is to compare generated (ideal) diffraction peaks with data from GIXD experiments. It allows to quick-check unit cell configurations and contact plane indices. Data must be loaded. Every unit cell input is tested for Niggli's criteria and the reduced cell according to Buerger is determined. If the criteria are met, the icon on the left-hand side ('Niggli's criteria') is switched to green. The type of the cell is indicated in the 'Information box' (*cf.* Figure 1.2d).
The lattice constants of the reduced cell are printed in the table on the right-hand side (*cf.* Figure 1.2c). If the input leads to a unit cell which can be reduced by the earlier introduced Buerger-algorithm, the newly derived parameters are printed. Otherwise, the input parameters are displayed at this location. As soon as the button 'Add crystal' is

pressed and the input is reviewed, the icon with the label 'Buerger cell' turns green. The button with the label 'Clear crystal' can be used to remove the input from the panel. All data points from the *Data representation* graph will be removed in this case.

Figure 1.2: (a) Screenshot of the indexing routine's *Add Crystal Panel*. (b) Image of the *Data representation* graph with loaded diffraction peaks (black circles) and generated peak positions (red crosses). (c) Table with parameters of the reduced cell. The table is cut out from (a) and scaled up to show all six lattice constants. (d) Icon with information box indicating the type of the cell according to Niggli's criteria.

The *Add Crystal Panel* is an additional feature and does not include any indexing of data points. A diffraction peak can be calculated in a straightforward manner, if all six lattice constants as well as the Miller and the Laue indices are known. With these input parameters, the ideal diffraction patterns are generated. **Please check if diffraction peaks are uploaded!** This is required due to internal data comparison. The generated peak positions are added to the *Data Points and Representation Panel*. If data points (diffraction peaks) are loaded, they can be easily compared with each other (*cf.* Figure 1.2b). Via 'Add crystal', the test algorithm for Niggli's criteria and the test for reduced cells

are applied. The calculated diffraction pattern is added to the graph. By clicking 'Clear table', the input numbers are cleared only. The 'Clear graph' button allows to delete data from the graph and at the same time keeping the input numbers at the table. The input 'Max. hk' limits the maximal number of the two Laue indices h and k. It is separated from the input variable 'Max. l' to allow more variability when testing generated patterns. Peak positions are calculated for permutations of Laue indices in the range from $(-Max.hk)$ to $(+Max.hk)$. The same applies for the maximal value of l. Plotting the generated diffraction points can help to confine the range of potential unit cell solutions. This *trial and error* approach is sometimes useful to figure out constraints in lengths and angles. These constraints can be applied in the subsequent indexing procedure, as the panel introduced in sub-chapter 1.3 allows to set restrictions to unit cell parameters and contact plane indices.

1.2 Data Points and Representation Panel

The first part of the indexing workflow includes upload of the diffraction data. The buttons to upload (and remove) peak positions are located within the *Data Points and Representation Panel* and labelled "Load data' and "Clear data', respectively. If the properties of the input data are fulfilled, peaks can be uploaded using the button 'Load data'. The data are printed in the table on the left-hand side (see Figure 1.3). Here, 'data point', 'peak' or 'peak position' always refers to a (q_{xy}, q_z)–tuple. Successfully uploaded peak positions are plotted directly in the *Data representation* graph. Most important, however, is that the file with the diffraction data should not contain any other data type than *numeric* floating-point numbers. A dot is used for decimal separation. When pressing the button to load data, a *Windows dialog box* appears and allows to select files of the type .xls or .xlsx. The data have to be provided as row vectors. The first column includes the q_{xy}–data, the second one the q_z–data. As this routine is designed to make use of the specular information, at least one entry in the first column has to be zero. An initialization program in the background searches for this data point. If at least one specular peak is detected, the icon labelled 'Specular scan' toggles to green. If no such peak is contained in the uploaded file, a message pops up printing 'No specular peak detected. Please reload data or use the function "Indexing without specular scan". The user should add the specular information to the file or consider to use the button 'Indexing without specular scan'. In this case, a sub-routine is started which does the indexing of the diffraction peaks without any specular information. Hitherto, this is a rather time-consuming procedure. The

inclusion of a specular scan is highly recommended. The procedure to derive unit cell parameter without a specular scan can be controlled with the program inputs listed in the panel ”Setting for indexing without specular scan”.

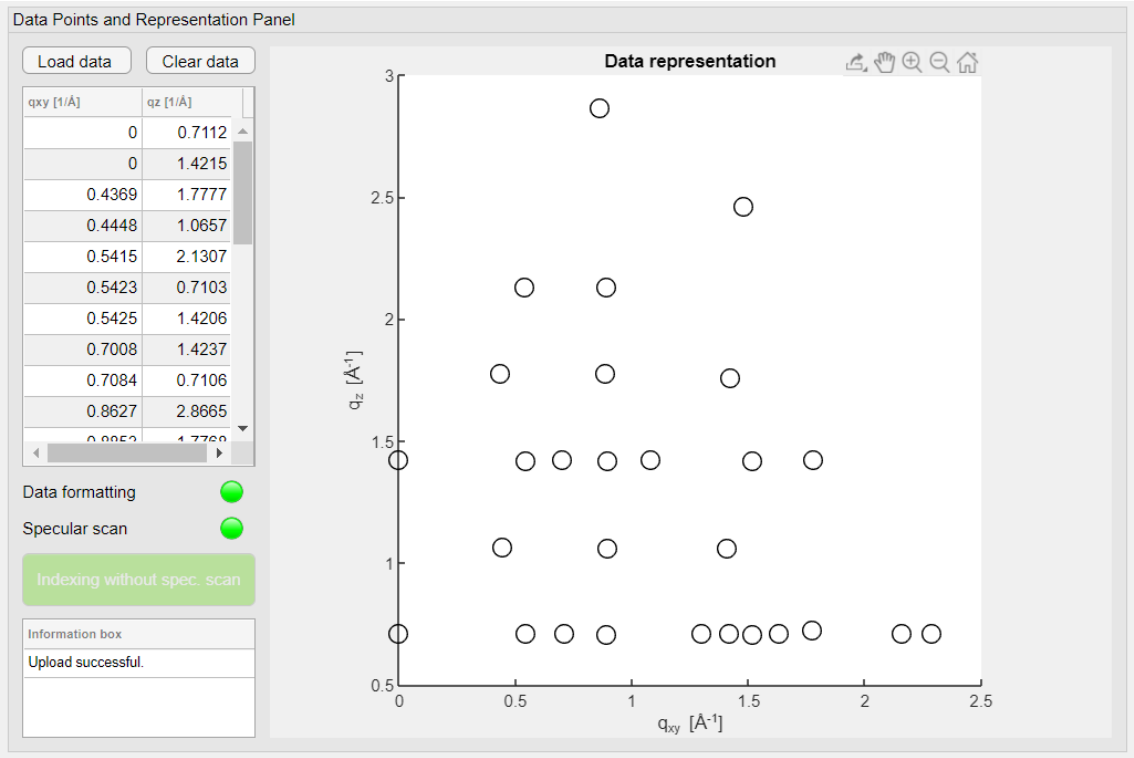


Figure 1.3: Screenshot of the indexing routine's *Data Points and Representation Panel*. An arbitrary set of peaks (black circles) is used to demonstrate the functionality.

The data set is not only checked for specular information, but also for the formatting. This includes the above mentioned representation as row vectors and the size of the data set (i.e. the number of data points). The icon labelled 'Data formatting' is switched to green if at least four (linearly independent) peaks are provided. This is the minimum number required to start this indexing routine. For data files with less than four peaks, the following message appears: 'Not enough peak positions for indexing. Please reload data!'. If the number of data points exceeds the value of 100, another message appears: 'The number of peak positions is above 100. Consider reducing the input'. To keep the computational effort in a reasonable range, this limit should not be exceeded in the beginning. If, however, enough restrictions (*cf.* Chapter 1.3) can be applied to the variables, the number of data points can be expanded without any constraint. After successful upload of the data points, the settings for the indexing algorithm have to be checked or adjusted.

1.3 Indexing Panel

The routine for indexing the uploaded GIXD data is controlled via the *Indexing Panel*. The region is marked with a green frame in Figure 1.1. Three tabs are used to control the routine and to provide maximal manual adjustability. The first tab, labelled *Indexing*, is used to review the input and to start, stop and reset the calculations. A screenshot is shown in 1.4. The button 'Check status of Part 1' triggers a subprogram to request certain parameters necessary for the routine. Such parameters are for instance the number of data points, contact plane settings (Miller indices), restrictions of lattice constants and presence of a specular peak. The panel interface right after start is shown in Figure 1.4a. Red icons indicate that no checks are requested. The buttons to control the routine are disabled in this case. If all conditions of the subprogram are fulfilled and 'answered', the upper icon turns green and further buttons are enabled, as shown in Figure 1.4b. Incomplete requests are indicated by popup messages of the kind 'Declined. Please load data and check formatting and specular scan!'.

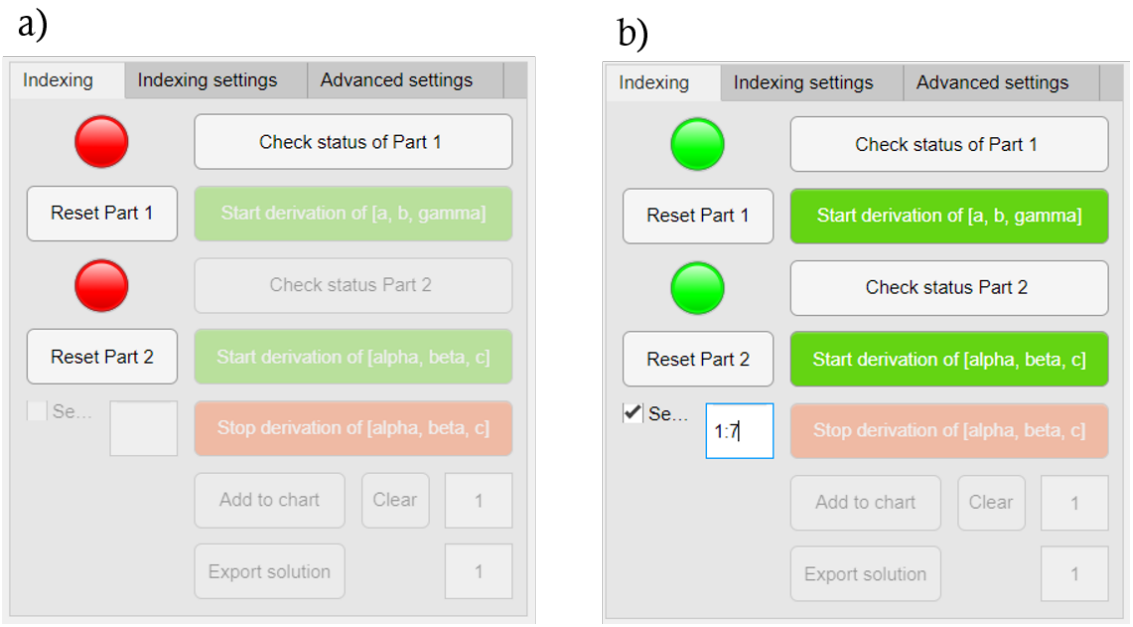


Figure 1.4: Screenshot of the routine's *Indexing Panel*. The image shows the tab 'Indexing' for two different states: (a) Right after start and (b) after derivation of (a, b, γ) and before continuing with final derivation of the lattice parameters.

The indexing routine can be started by pressing the button 'Start indexing Part 1'. After indexing the data in q_{xy}, the results for the (a, b, γ)-sets are shown in the *Result Panel*. The status for the second part has to be checked by pressing 'Check status Part 2'. If both icons are set to green, the second part of the indexing routine can be started by pressing

the button 'Start indexing Part 2'. The buttons 'Reset Part 1' and 'Reset' part 2' allow to set the routine to initial state (*cf.* Figure 1.4a).

Before starting the indexing procedure, certain adjustments can be done using the tabs *Indexing settings* and *Advanced setting*. In any case, the settings should be reviewed after successful upload of the peak positions. As mentioned earlier, applying constraints to the unit cell parameters can help to reduce computational effort. This can be done in the tab labelled as *Indexing settings*. The range for every of the six lattice constants as well as the corresponding cell volume can be limited by tick off the according checkbox placed on the left-hand side of the panel (*cf.* Figure 1.5). The 'Contact plane settings' are located at the bottom of the panel. Three modes are provided. If the first checkbox is enabled, the Miller indices are set to the values $u = 0$, $v = 0$. The indexing program for the special case is applied in this case. The third index, w, is varied in the range from $(-Max.Millerindex)$ to $(+Max.Millerindex)$. This setting can be changed in the *Advanced settings* tab, row 5. Only one checkbox can be chosen at a time. The second checkbox allows to define one specific set of indices, without any permutations. The indices have to be separated by using a comma in the form u,v,w. The input has to be confirmed by pressing the ENTER key and is included to the algorithm by pressing 'Apply changed settings'. Changes are only accepted after this button is pressed. The messages in the 'Information box' inform about the current state of the settings. All input can be set to default settings by pressing the button 'Reset settings'.

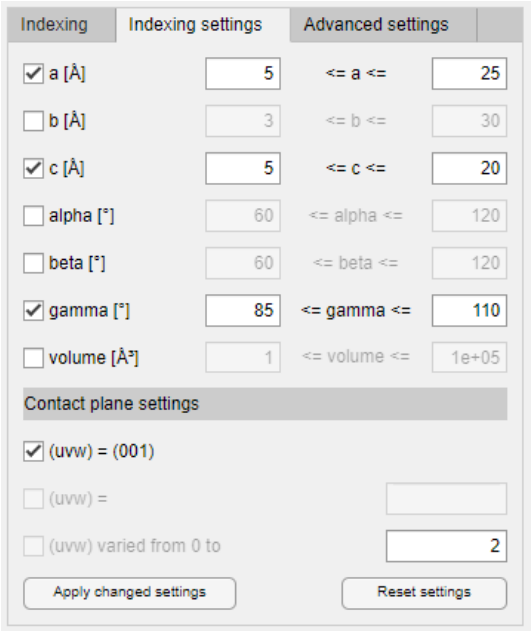


Figure 1.5: Screenshot of the routine's *Indexing Panel*. The image shows the tab labelled as *Indexing settings*. For demonstration purposes, three unit cell parameters are restricted and the 'Contact plane settings' are adjusted for the special case.

The last checkbox allows to define a permutation range for the three Miller indices of the contact plane. In Figure 1.5, the input is set to a value of 2. In this case, every possible permutation of (u,v,w) in the range from -2 to $+2$ is generated and tested by the indexing program. The limit for the permutations can be set in the tab labelled as *Advanced settings* via the variable 'Max. Miller index'. After the variables are reviewed and/or adjusted in the *Indexing settings* tab, the routine is ready to use. Further tuning of certain variables is possible in the tab labelled as *Advanced settings*, however, not necessarily required. This tab requires some understanding of the algorithm. It may occur that no unit cell solutions can be obtained by use of the default settings or that only solutions with unsatisfying errors in the components of $\mathbf{q}$ are derived. In this case, re-adjusting of the program parameters in the *Advanced settings* tab can be useful. Even if solutions are obtained, the sets can be refined by the use of these settings (i.e. by change of maximum order of the Laue indices). The tab contains ten parameters (*cf.* Figure 1.6). Determination of potential unit cell solutions requires solutions to linear systems of equations (LSEs).

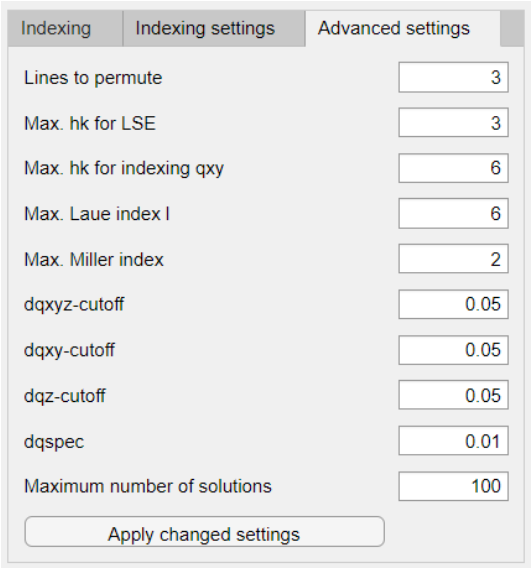


Figure 1.6: Screenshot of the routine's *Indexing Panel*. The image shows the *Advanced settings* tab. The settings should not be changed, unless enough restrictions to the unit cell parameters can be applied.

The start set of this system needs three linearly independent pairs of (q_{xy}, q_z). If the parameter 'Lines to permute' is set to the value of 3, a subprogram checks for linear dependency and sets up LSEs for these three values with every permutation in the Laue indices between $(-Max.hk\,for\,LSE)$ to $(+Max.hk\,for\,LSE)$. With increase of the parameter 'Lines to permute' more combinations of given q-values are considered what makes it more likely to determine solutions. After possible solutions are provided through solving the LSEs, the in-plane components of **q** are indexed. This occurs by assignment of the Laue indices and subsequent evaluation of the RMSD (root-mean-square-deviation). The program parameter 'Max. hk for indexing qxy' allows to specify the upper limit for this indexing. In a refinement run (i.e. after the range of the lattice constants can be restricted), this value can be increased to a value of 8 or more. It is advised not to change this parameter in the beginning. The parameter 'Max. Laue index l' defines the limit for the second set of the LSEs. As discussed earlier for the *Indexing settings*, the maximal range for the Miller index permutations is defined by the parameter 'Max. Miller index'. The next program parameters of the *Advanced settings* tab relate to the output of the determined solutions. The table where the results are shown (marked in yellow in Figure 1.1) contains only solutions where the error in q_{xyz}, q_{xy} and q_z is lower or equal to 'dqxyz-cutoff', 'dqxy-cutoff' and 'dqz-cutoff', respectively. Furthermore, no solutions are indexed if the error in the specular peak is greater than 'dqspec'. If the routine is started and not stopped manually by the user, the algorithm will stop automatically

as soon as the value 'Maximum number of solutions' is reached. All changes have to be confirmed by pressing 'Apply changed settings'. The status is then indicated in the 'Information box'.

1.4 Error Panel

During the indexing procedure, the root mean square deviation (RMSD) is calculated for the components q_{xy} and q_z as well as for the length q_{xyz} of the scattering vector. The length can be obtained from the experimental data by $q_{xyz} = \sqrt{q_{xy}^2 + q_z^2}$. The error in the length turned out to be a promising indicator for the choice of a possible unit cell solution and it is therefore used as a sorting quantity. For this reason, the error is displayed throughout the indexing procedure over the number of sweeps (*cf.* Figure 1.7).

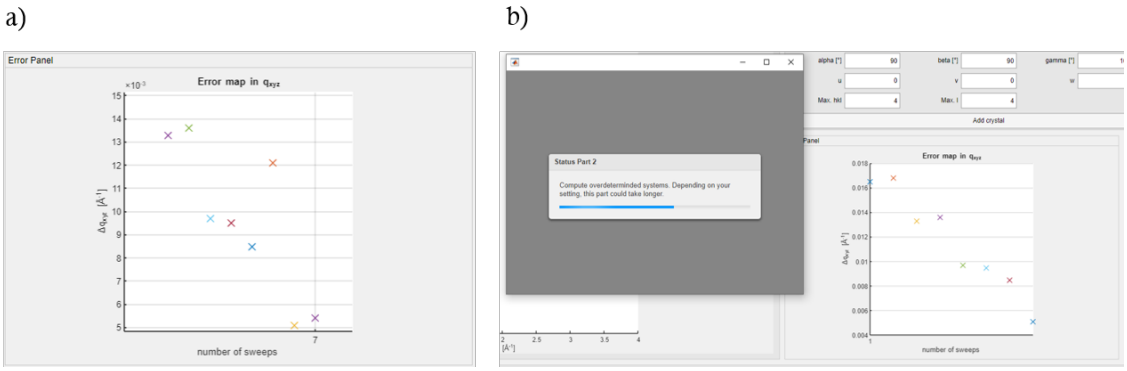


Figure 1.7: Screenshot of the indexing routine's *Error Panel*. In (a) the image shows the panel as embedded in the graphical user interface. The image in (b) shows a cutout from the main window while the routine is running. The error decreases with every sweep and the routine is stopped after the error is below 0.6 %.

One sweep contains the obtained solutions from one set of (a, b, γ). The minimal error of these solutions is printed in the graph labelled as 'Error map in q_{xyz}'. The indication can be used to monitor the error and to stop the algorithm if a proper value is met.

1.5 Result Panel

The unit cell solutions obtained from the indexing routine are presented in the form of two tables. These tables are located in the *Result Panel* and marked with a yellow frame in Figure 1.1. To give a better overview, the panel is divided in two and shown separately in Figure 1.8. Every row corresponds to one possible solution. The solutions obtained

from the first part are stored in the left table (*cf.* Figure 1.8a) and labelled with a number (#). The output is sortable by every quantity shown in the table by pressing the ↑↓-icon, right beside to the variable names. In addition, the sets can be sorted by smallest area by pressing 'Sort by parallelogram area'. The default sorting can be restored with the button 'Sort by dqxy'. As the first part of the algorithm is based on parallel-executed loops, it can be stopped only during the initialisation phase. Usually this takes up to 20 seconds after start up. To this point, the program can be stopped by pressing 'Stop Part 1'.

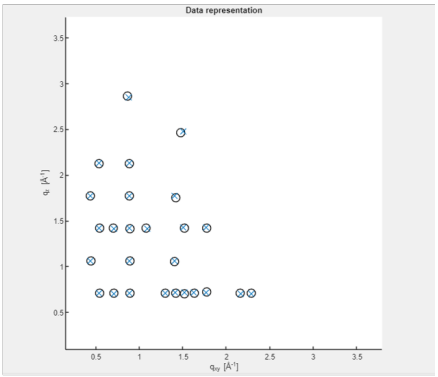
a)

Result Panel

#	u	v	qspec [...	a [Å]	b [Å]	gamma...	dqxy [1/Å]
74.0000	0	-1.0000	0.7112	14.0866	14.5608	90.7164	0.0196
75.0000	0	1.0000	0.7112	14.0866	14.5608	90.7164	0.0196
76.0000	0	-1.0000	0.7112	14.0873	14.5843	89.1428	0.0193
77.0000	0	1.0000	0.7112	14.0873	14.5843	89.1428	0.0193
78.0000	0	0	0.7112	11.5870	17.7383	89.9973	0.0251
79.0000	0	0	0.7112	8.8691	23.1740	89.9973	0.0288
80.0000	0	0	0.7112	14.4845	14.7017	74.8620	0.0077
81.0000	0	0	0.7112	14.4845	14.7017	105.1380	0.0077
82.0000	1.0000	-1.0000	0.7112	12.5144	16.9678	90.1806	0.0133
83.0000	-1.0000	1.0000	0.7112	12.5144	16.9678	90.1806	0.0133
84.0000	-1.0000	-1.0000	0.7112	12.5142	17.0289	89.9206	0.0134

Stop Part 1 | Sort by parallelogram area | Sort by dqxy

c)



b)

#	u	v	w	a [Å]	b [Å]	c [Å]	alpha [°]	beta [°]	gamma [°]	volume [Å^3]	dqxyz [1/Å]	dqz [1/Å]	dqxy [...	dqspec [1/Å]
54	0	0	2	14.5372	14.6549	17.7390	90.3691	90.1334	104.8954	3.6520e+03	0.0062	0.0067	0.0084	0.0043
55	0	0	2	14.5155	14.6711	17.7750	90.4185	89.8276	75.0913	3.6578e+03	0.0062	0.0071	0.0086	0.0066
56	0	0	2	14.5188	14.6937	17.6448	90.4240	89.9383	75.3712	3.6421e+03	0.0062	0.0076	0.0088	0.0024
57	0	0	2	14.5123	14.6747	17.7785	90.4386	90.1571	104.8541	3.6595e+03	0.0062	0.0071	0.0085	0.0066
58	0	0	2	14.5197	14.7057	17.6680	89.6800	89.9962	74.8821	3.6419e+03	0.0062	0.0073	0.0087	3.5834e-04
59	0	0	2	14.5286	14.7004	17.6711	89.7291	90.0530	74.8658	3.6432e+03	0.0062	0.0071	0.0091	2.8270e-04
60	0	0	2	14.5140	14.7083	17.6708	89.6858	90.0260	105.0585	3.6427e+03	0.0062	0.0072	0.0085	3.5834e-04
61	0	0	2	14.5120	14.7066	17.6543	90.2178	89.7406	74.8974	3.6376e+03	0.0062	0.0071	0.0087	0.0015
62	0	0	2	14.5200	14.6678	17.7142	90.3615	90.1514	104.6062	3.6507e+03	0.0063	0.0061	0.0086	0.0023
63	0	0	2	14.5443	14.6825	17.6522	89.6743	90.0496	75.2833	3.6458e+03	0.0063	0.0073	0.0087	0.0020
64	0	0	2	14.5550	14.6975	17.6953	89.7491	90.1075	74.9384	3.6553e+03	0.0063	0.0070	0.0082	8.4961e-04
65	0	0	2	14.5308	14.6602	17.7431	89.6127	90.1213	75.1293	3.6530e+03	0.0063	0.0066	0.0082	0.0049
66	0	0	2	14.5060	14.7299	17.6318	90.3538	90.2923	105.0608	3.6378e+03	0.0063	0.0076	0.0194	0.0028

Figure 1.8: Screenshots of the indexing routine's *Result Panel*. (a) shows the table containing the results from the first indexing part. The complete unit cell solutions obtained from the indexing routine are presented in the table shown in (b). The image in (c) shows a cutout from the *Data Points ans Representation Panel*. The peak positions calculated from one solution (blue) are printed together with the experimental data points (black).

The final results after the second indexing part are presented in the table located at the right-hand side. It is shown as cutout in Figure 1.8b. The solutions are labelled with numbers and sorted by the RMSD in q_{xyz}. Again, all solutions can be sorted with the ↑↓-icon. For demonstration purposes, the solution with row number 58 is selected and added to the *Data representation* graph (*cf.* Figure 1.8c).

1.6 Tutorial

1.6.1 Indexing using a specular peak

The application of the program is presented in terms of a short tutorial on a organic thin film sample (”Cu INA”- MOF). A set of 26 peak positions is extracted from GIXD patterns using the software tool GIDVis [1]. Two additional data points are obtained from XRD experiments measured in specular condition. This leads to a total of 28 peaks for the indexing procedure and for subsequent determination of a unit cell solution.

Having the MATLAB executable (here ’GIDInd_2021_02_25.exe’) installed, the app can be started rightaway. The graphical user interface starts, expands to full-screen size and is then compressed to a default size. The icons may occur compressed or distorted, depending on the screen size. The GIXD data are uploaded by pressing the ’Load data’ button and selecting the according file. The file has to be of the format .xls or .xlsx. It is possible that the main window is minimized during the selection. The window expands again as soon as a file is selected. The uploaded data points are added directly to the graph in the form of black circles. As all formatting requirements are fulfilled and two specular peaks are contained in the data, both icons are set to green. Next, the indexing settings are checked. As there’s no prior information, the routine is operated in general mode to estimate the range of potential unit cell constants. The program parameter ’Max. Miller index’ is set to 1. The Miller indices are therefore of the kind $(11w)$, $(10w)$, $(00w)$ and so on. By pressing the button labelled as ’Apply changed settings’, the changes are accepted. The indexing procedure is initiated by checking the settings one more time by pressing ’Check status of Part 1’. A green lamp and enabled buttons indicate that the program is ready. The routine is started by pressing ’Start derivation of [a,b,gamma]’ and ’Proceed’ in the subsequently opened dialog box. After start up, a window containing status messages opens. This window informs the user about the current state of the algorithm. The very first run of the routine may takes more time due to startup of the MATLAB parallel pool.

The first part of the indexing routine is finished as soon as results appear in the left table of the *Result Panel*. To continue indexing, the settings have to be checked by pressing 'Check status of Part 2'. Two options are enabled for the further process: (i) start indexing the sets of (a, b, γ) in ascending order or (ii) ticking the checkbox and define specific lines to consider for the second indexing part. The checkbox is located below the button 'Reset Part 2'. The rows have to be specified using MATLAB syntax. The screenshot in Figure 1.9 shows the indexing routine in use. In this particular case, the lines are not furter specified.

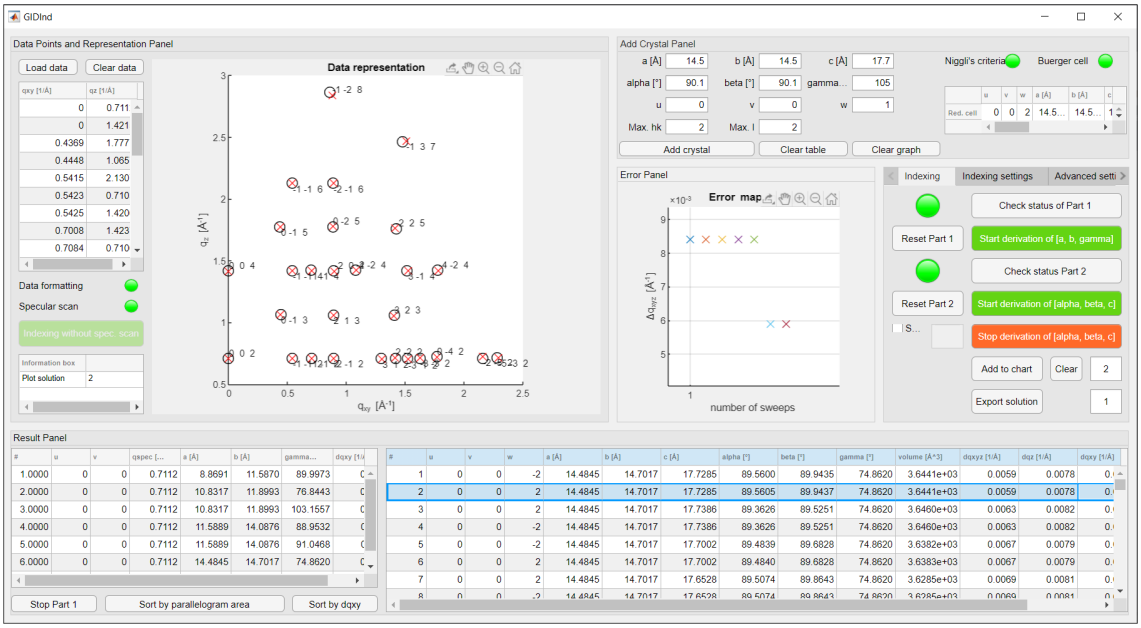


Figure 1.9: Screenshot of the indexing routine's *Main Window* in use. The indexing procedure is finished and the peak positions obtained from the solution highlighted in blue is added to the data representation graph.

By pressing the button 'Start derivation of [alpha, beta, c]', the routine assigns the remaining indices and calculates the lattice constants α, β and c. In this example the routine is stopped after seven sweeps as the error in q_{xyz} is in a range of below 1 %. With an error in this range, the solution can be 'tested' by adding the calculated diffraction pattern to the graph. The line number has to be entered in the field beside the 'Add to chart' button. After confirming by pressing the ENTER key and pressing the button 'Add to chart', the generated peak positions are added to the graph. The peak positions generated with this unit cell solution show good agreement with the experimental data. For further comparison, the solution can be used as input for GIDVis. By use of the according Laue indices, the diffraction peaks can be plotted directly on top of a reciprocal space map (*cf.* Figure 1.10). The Laue indices are provided by the indexing routine in the form of a .xls-

file. The file can be generated by input of the row number in the corresponding field and pressing the button 'Export solution file'. An example of such a file is shown in Figure 1.11. It is advised to use the software GIDVis together with the graphical comparison of the indexing routine to achieve the best results.

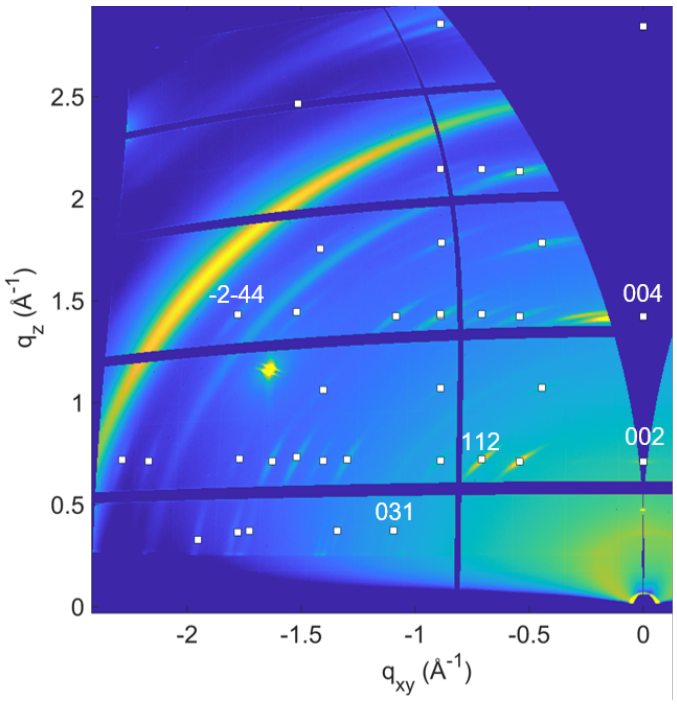


Figure 1.10: Screenshot of a reciprocal space map generated in GIDVis. The white squares represent the calculated peak positions using the derived unit cell solution and the Laue indices provided by the indexing routine. Five data points are presented with the according Laue indices.

The errors in length q_{xyz} and in the two components of the scattering vector are helpful criteria for the search of a unit cell solution. However, the final choice of such a solution depends also on other factors (e.g. highest symmetry of the cell and shortest lattice edge-lengths). The here shown unit cell solution is not the only possible solution. For the sake of a compact tutorial however, the discussion of other possibilities is postponed to Chapter 2.

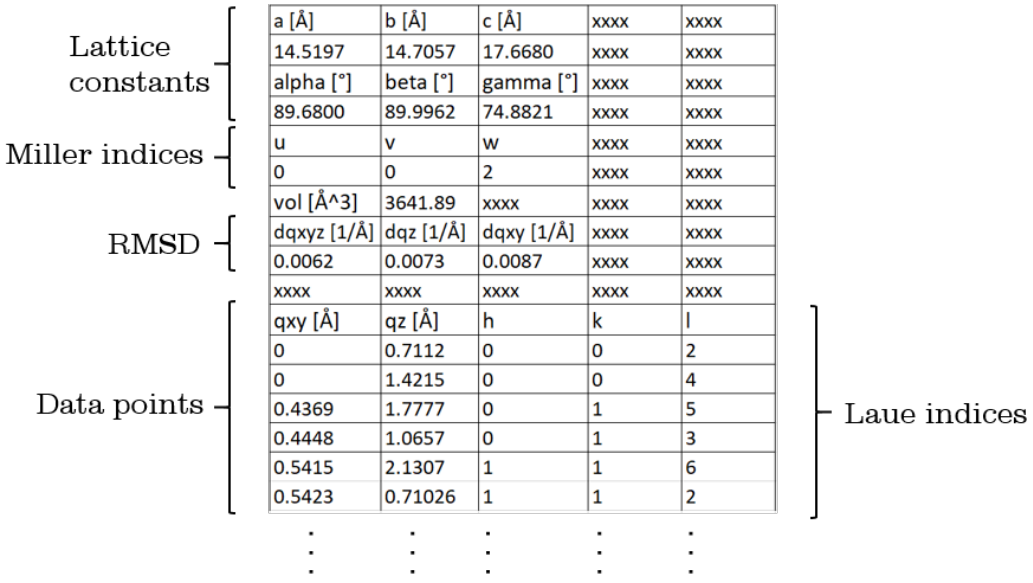


Figure 1.11: Screenshot of the output file generated by the routine. For the sake of brevity, the file is cut after six peak positions. The locations of the different output information are marked separately with braces. If the sub-routine 'Indexing withou spec. scan' is used, two additional parameters are stored in the first two rows.

The mathematically optimized solution chosen for this short tutorial is a cell of type I with the following unit cell parameters:
$u = 0$, $v = 0$, $w = 2$, $a = 14.48$ Å, $b = 14.70$ Å, $c = 17.73$ Å, $\alpha = 89.56°$, $\beta = 89.94°$ and $\gamma = 74.86°$. The volume of the reduced Buerger cell is $V = 3644$ Å^3 and the mean deviation in q_{xyz} is 0.6 %. An additional check using the *Add Crystal Panel* of the indexing routine confirms that the solution obeys the scalar-product criteria according to Niggli.

1.6.2 Indexing without specular scan

The indexing routine GIDInd allows derivation of the lattice parameters not only by including the specular scan. A sub-routine offers the possibility to try to derive solutions without this specular peak at $q_{xy} = 0$. It should be emphasized that this option is an add-on which is not fully tested. The main focus of the GIDInd indexing software is by including a specular peak in the GIXD diffraction data. Although not fully tested, three data sets are indexed and their corresponding unit-cell solutions were derived with this sub-routine. Please note that the latest stable version offering 'Indexing without specular scan' is the executable file 'GIDInd_2021_02_25.exe'. Again, the peak positions are the required input data for the here proposed routine GIDInd. Using the GIXD data only, this approach is a single-step procedure with every system solved yielding one potential

set of lattice constants. In a subsequent step, every unit-cell solution must be rated with respect to the deviation from the measured diffraction pattern. Due to the high number of possible combinations of the three Laue indices, this can be computationally challenging often limited by performance and memory of the available machine. The presence of a specular scan provides data in a region which is not accessible with conventional GIXD experiments. This additional peak allows to split the derivation of the lattice constants into two consecutive parts. In this way, it is not necessary to compute and evaluate an entire set of lattice parameters for every single combination of integer numbers which constitute the Laue indices of diffraction. It is rather possible to intervene at an earlier step and to sort and filter solutions, thereby keeping the computational effort at a lower level. Moreover, the orientation of the crystallographic cell on the substrate surface is explained by the Miller indices of the plane parallel to the substrate surface. Subsequently, the work flow for derivation of unit cell solutions without a specular scan is elaborated. For the algorithm to start, at least six peak positions are required to derive a triclinic cell. To start the routine with a monoclinic cell, a minimum of four peak positions must be provided. However, this is the lower limit and it is recommended to use more peaks for the indexing procedure. When loading a data set without a specular peak, the program will show a dialog box with the message 'No specular peak detected. Please reload data or use the function 'Indexing without specular scan". After confirming by pressing 'Ok', the main window appears again with the now enabled green button 'Indexing without spec. scan'. Press this button. A dialog box as shown in Fig. 1.12 appears. All possible settings for the sub-routine are summarized within this panel. The main window does not contain settings for indexing without a specular scan included. The user is able to choose between two programs. The first one assumes a monoclinic cell where the cell angles α and γ are fixed at 90°. The third angle, β, is free. Restrictions on the free angle as well as on the remaining cell parameters and the volume can be given using the check boxes. Please confirm any changes to the lattice parameter and also the change from 'monoclinic' to 'triclinic' option by pressing the green button 'Refresh'.

The indexing settings are shown on the right-hand side of the panel. 'For calculation of the lattice parameters' refers to solve the linear systems of equations. By solving these, the six lattice parameters are all derived at once. The maximum value for the Laue indices are limited due to potential computational overload. As mentioned earlier, this is just an additional option and not recommended for general indexing of the GIXD data. The variable 'Lines to permute' defines how many start values are takes from the uploaded data set to derive potential solutions for the lattice constants. The value must not

be set lower than 4 in the monoclinic and 6 in the triclinic case.

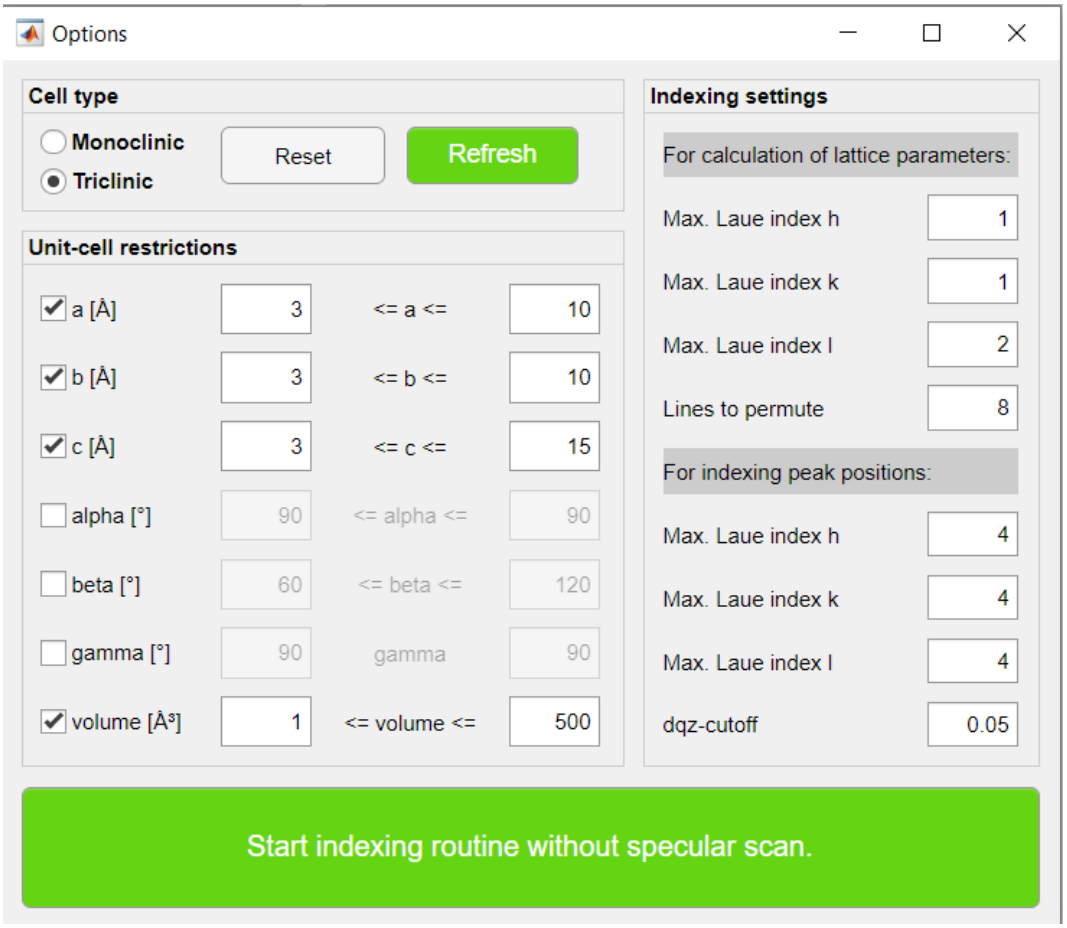


Figure 1.12: Screenshot of the dialog box for the sub-routine 'Indexing without specular scan'.

The value 8 has shown to be sufficient for at least three different and independent samples. Under the tab 'For indexing peak positions' the maximal Laue indices can be chosen. The limit 'dqz-cutoff' is a numerical parameter to narrow the number of obtained solutions. In general, 10 % is a good start value. This corresponds to the input value 0.1. After all parameters are set, the routine can be started by pressing 'Start indexing routine without specular scan'. The panel does not close until the calculations have finished. The information box prints the status of the routine in this sub-routine. As long as the message 'Busy. Searching monoclinic/triclinic cell' is printed, the routine is running. Finished derivations are indicated by closing the panel and printing 'Finished. Please check Result panel' into the information box. As soon as the solutions appear within the 'Result Panel', plotting and export is possible in the same way as stated earlier. Note that the contact plane can not be given in terms of integer Miller indices. When exporting the solution and saving it as .xlsx-file, two additional parameters are stored. Namely the rotational angles Φ and Ψ. For usage and further information of these parameters please

read [2, 3, 4].

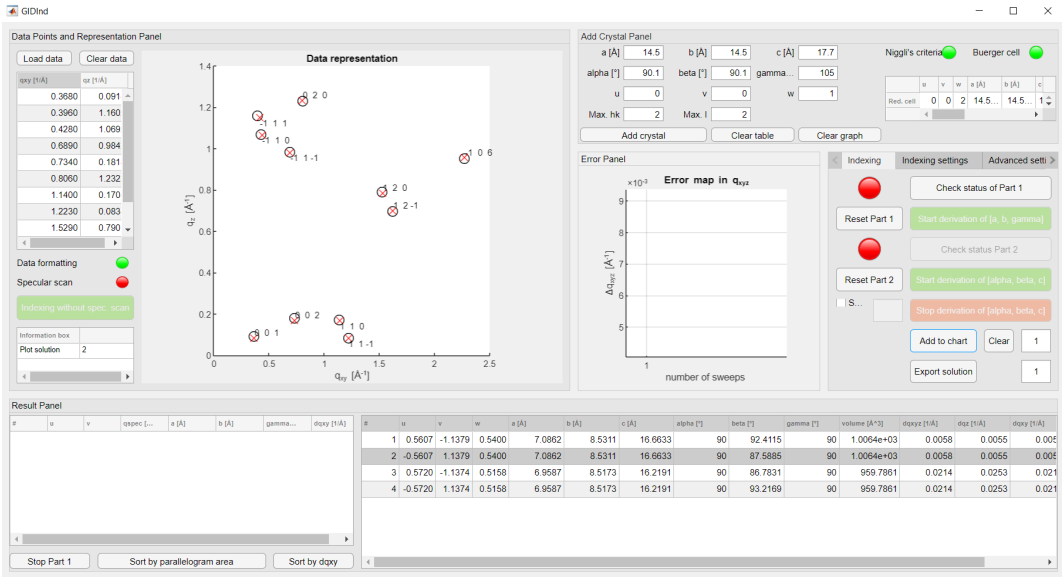


Figure 1.13: Screenshot of the main window after using 'Indexing without specular scan' for DIP. Black circles mark the input data and red crossed refer to the simulated diffraction pattern using the as-derived solution 1 with $a = 7.09$ Å, $b = 8.53$ Å, $c = 16.66$ Å, $\alpha = 90.00°$, $\beta = 92.41°$ and $\gamma = 90.00°$. The volume of the reduced Buerger cell is $V = 1064$ Å^3 and the mean deviation in q_{xyz} is appr. 0.6 %.

After finished derivation, the main window can be used as in the beginning. The screenshot in Fig.1.13 shows the GUI after derivation of the lattice constants for a diindenoperylene (DIP) thin film sample grown on pyrolytic graphite. The same data set (but with specular scan) is investigated in the next chapter. In total, the functionality of this subroutine has been tested for three different and independent data sets. The sets are contained in the zip-file online and saved as 'dataset_PQ_75_peaks' and 'dataset_aspirin'. For the latter one, please remove the specular peaks manually. The file used to demonstrate the functionality here is named 'dataset_DIP_without_qspec'. For further information please visit `https://www.if.tugraz.at/amd/GIDInd/` and `https://github.com/m-kainz/GIDInd`.

Chapter 2

Indexed Samples

The Indexing Routine is used for indexing and determination of the unit cell solution of four different thin film samples. One example, namely the Cu INA - MOF fIna-04, was shown earlier in sub-chapter 1.6. In this final chapter, the program is demonstrated on three further samples. The input data were derived using GIDVis and provided as .xls-files. The used data are explicitly printed in the Appendix A. Results are compared with published values from literature if available. Note that these examples are derived using the GIDInd version 'GIDInd_2021_01_14.

2.1 Diindenoperylene (DIP)

Grazing incidence X-ray diffraction data of a DIP thin film on highly oriented pyrolytic graphite (HOPG) are evaluated in this sub-chapter. Diindenoperylene is an organic semiconductor molecule with the chemical formula $C_{32}H_{16}$. A specular peak is provided at $q_z = q_{\text{spec}} = 1.776$ Å^{-1} and used together with 11 other diffraction peaks for the subsequent indexing process. The routine is started with default contact plane settings where the Miller indices are varied in a range between -2 and $+2$. A total number of $5^2 = 25$ different combinations result which are tested in parallel manner. The partial solutions from the first indexing part are sorted by the RMSD in q_{xy} is ascending order. The second indexing part is cancelled after 30 sweeps as the error in q_{xyz} does not decrease with increasing number of possible sets. The routine is re-started with readjusted settings where the number of start sets is increased by setting 'Lines to permute = 5' and 'Max. hk for LSE = 4'. The contact plane settings are unchanged and the program runs again through 25 different pairs of Miller indices. The second indexing block yields first promising results after approximately 20 sweeps and 10 minutes. The errors are in the range of 1

% and below, from this point on. A screenshot from this state of indexing is shown in Figure 2.1. The solution with unit cell parameters $a = 8.49$ Å, $b = 14.27$ Å, $c = 16.64$ Å, $\alpha = 87.1°$, $\beta = 89.2°$ and $\gamma = 89.9°$ is plotted in the *Data representation* graph. The corresponding derived contact plane is indicated by (uvw) =(-2-21). The errors can be explicitly given with $\Delta_{q_{xy}} = 0.006$ Å^{-1}, $\Delta_{q_z} = 0.001$ Å^{-1} and $\Delta_{q_{xyz}} = 0.004$ Å^{-1}. The numerical deviations are in a low region and as the calculated pattern does practically not depart from the experimental data, this solution is chosen for a further investigation. In a next step, the program is used to test specific contact plane variations based on the numbers derived above. Moreover, numerical restriction are made to concentrate the calculation on solutions close to the above determined solution range.

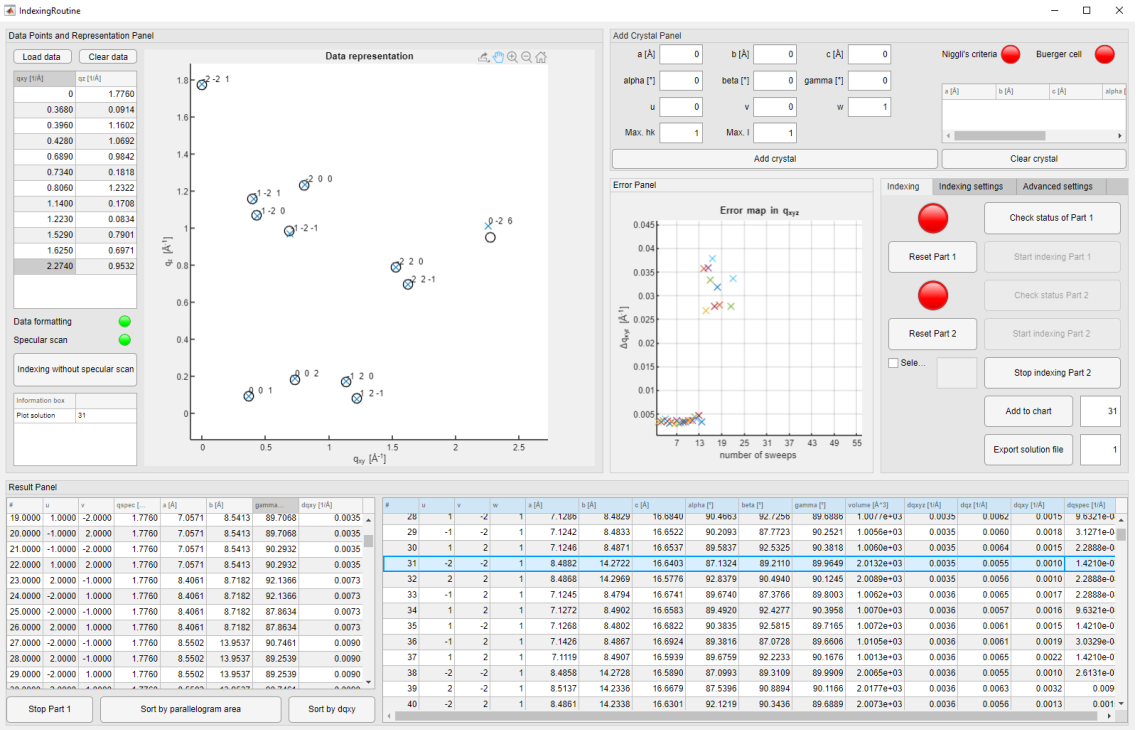


Figure 2.1: Screenshot of the main window of the Indexing Routine upon indexing of the provided DIP data. The solution from the *Result Panel*, marked in blue, is used to generate a simulated diffraction pattern. The pattern is added to the *Data Points and Representation Panel* for graphical comparison with the experimental data points.

After cycling through different permutations of the Miller indices consisting of combinations like (221), (2-12) or (121), a concluding mathematical solution can be recommended. The Miller indices can be stated as (uvw) =(-121) and the corresponding lattice constants are $a = 7.13$ Å, $b = 8.48$ Å, $c = 16.67$ Å, $\alpha = 89.4°$, $\beta = 87.8°$ and $\gamma = 89.7°$. The cell volume is derived with $V = 1070.1$ Å^3 and the mean deviations for this solution

are $\Delta_{q_{xy}} = 0.002$ Å^{-1}, $\Delta_{q_z} = 0.005$ Å^{-1} and $\Delta_{q_{xyz}} = 0.002$ Å^{-1}. Note that a very similar solution is contained two rows below the highlighted solution in Figure 2.1 above. Indexing and the resulting solutions of the DIP sample can be used as a vivid example for the interchangeability of the Miller indices and the determined unit cell parameters. From a mathematical point of view, every symmetric solution should be contained and provide an equivalent description of the crystallographic unit cell. This effect was observed for the defined Miller and Laue indices at the here shown example and is reported in literature [2]. The indexing procedure of this particular sample shows a different approach, compared to the example of Naproxen. Here, a promising solution is derived and investigated further by gradually limiting the cell parameters under testing of different Miller indices. Upon gradual application of the routine and its features, a reduced form of the preliminarily derived cell can be determined. A similar unit cell solution was reported earlier in literature [3, 5].

2.2 Pentacenequinone (PQ)

A total number of 30 diffraction peaks is used for indexing the GIXD pattern of a PQ thin film sample. 6,13-pentacenequinone is another example of an organic semiconductor material with the chemical formula $C_{22}H_{12}O_2$. The specular peak is derived from XRD and located at $q_z = 1.946$ Å^{-1}. For indexing of this sample, no changes at all are made in the first approach. The data set is uploaded and the indexing procedure is started right away with the default settings. The obtained sets of (u, v, a, b, γ) are processed without any further sorting. The routine is stopped manually after sweep 31. A screenshot of this state is shown in Figure 2.2 and the best ranked result (#1) can be explicitly stated as $u = 1$, $v = 0$, $w = 2$, $a = 5.06$ Å, $b = 8.08$ Å, $c = 8.87$ Å, $\alpha = 91.5°$, $\beta = 93.2°$ and $\gamma = 94.2°$. This is the reduced solution of the unit cell. The volume of this cell is $V = 360.8$ Å^3 and the mean deviation in q_{xyz} is 0.003 Å^{-1}. The RMSDs with regard to the components of the scattering vector are $\Delta_{q_{xy}} = 0.003$ Å^{-1} and $\Delta_{q_z} = 0.01$ Å^{-1}. As mentioned already for the DIP sample, mathematically symmetric solutions appear for every of the stated sets.

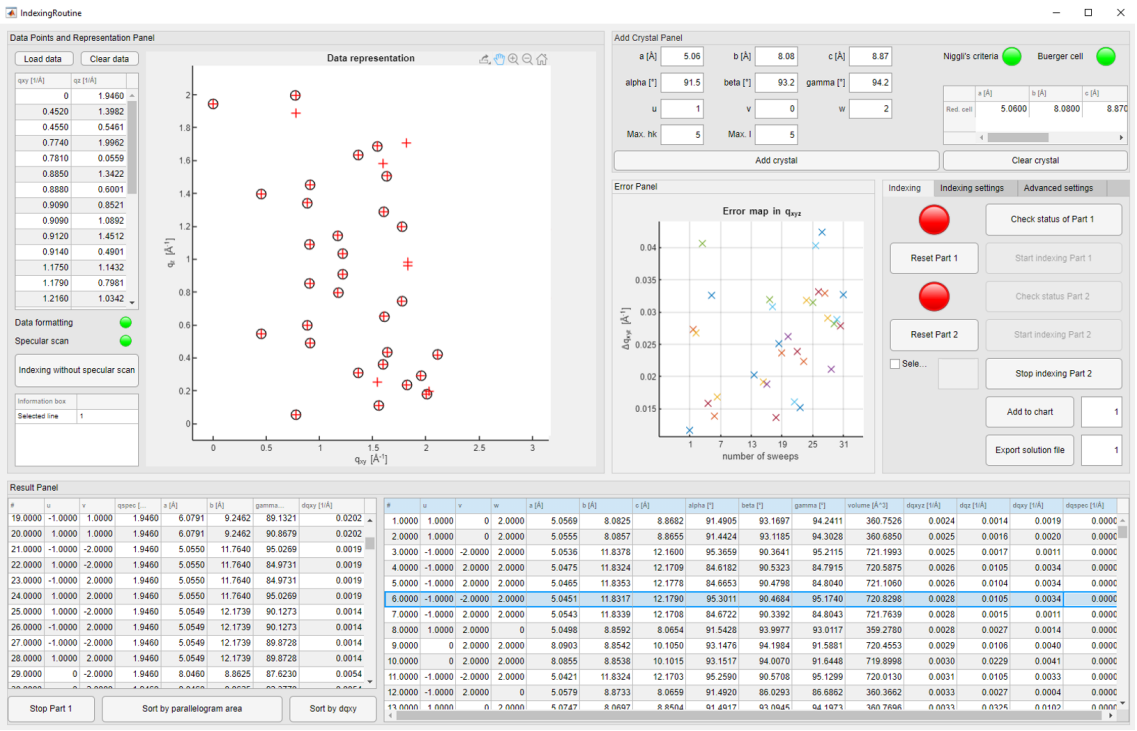


Figure 2.2: Screenshot of the Indexing Routine upon indexing of the provided PQ data. The two obtained solutions are shown within the *Result Panel*: The first one is ranked nr. 1 and the other solution is highlighted in blue. The experimental peak positions are printed as black circles and the the red crosses represent the generated diffraction pattern by applying the *Add Crystal Panel* for the first solution. Note that there are more data points generated as provided for indexing. For reasons of overview, generated peaks beyond $q_{xy} = 2.5$ Å^{-1} are deleted.

The second unit cell solution is explicitly given by the edge lengths $a = 5.05$ Å, $b = 11.83$ Å, $c = 12.18$ Å and the three angles are Å , $\alpha = 95.3°$, $\beta = 90.5°$ and $\gamma = 95.2°$. With a volume of $V = 720.8$ Å^3, the cell is approximately twice as large as the reduced cell stated above. The contact plane is derived with (uvw) =(-1-22) The deviations are in the same range with $\Delta_{q_{xy}} = 0.003$ Å^{-1}, $\Delta_{q_z} = 0.01$ Å^{-1} and $\Delta_{q_{xyz}} = 0.003$ Å^{-1}. Both unit cell solutions were found in literature [2]. Anyhow, a final comparison using the original GIXD data and GIDVis is recommended.

2.3 Naproxen

Diffraction data from a thin film sample of the organic molecule Naproxen are provided for indexing to investigate the crystalline phase. Naproxen is a pharmaceutical active ingredient and used for nonsteroidal anti-inflammatory drugs. A total of 30 peak positions are used for determination of a unit cell solution. The specular peak is obtained

from XRD measurements and is already included in the data set. In the first approach, the Indexing Routine is operated with contact plane settings $(uvw) = (001)$. No changes are done in the 'Advanced settings' and for the time being, no restrictions are applied to the unit cell constants. The first indexing part takes less than five minutes and approximately 130 partial solutions can be determined with an error smaller than 1 % in q_{xy}. With sorting the results according to the error in q_{xy}, the second part of the indexing procedure is pursued. The obtained unit cell solutions from the *Results Panel* are plotted and compared with the experimental peak positions using the *Data Points and Representation Panel*.

As the graphical comparison did not yield any comprehensive solution with the chosen settings, the routine is restarted. This time, the initial partial results are sorted regarding the smallest area of the parallelogram. Additionally, the program parameter 'Lines to permute' as well as the parameter 'Max. hk for LSE' are increased to a value of 4. The routine is stopped after 50 sweeps as the root mean square error, which is printed in the *Error Panel*, fell below 1 % in q_{xyz} a number of times. The unit cell constants a, b and c accumulate in regions around 10 Å, 20 Å and 25 Å, respectively. Similar patterns and repetitive behaviour is observed for the angles α, β and γ. The obtained solutions were added to the *Data representation* graph one after another for graphical comparison (*cf.* Figure 2.3). In addition, similar unit cell solutions were tested using the *Add Crystal Panel*. Indexing of the data set using general contact plane settings yields symmetric solutions with regard to the ones stated above.

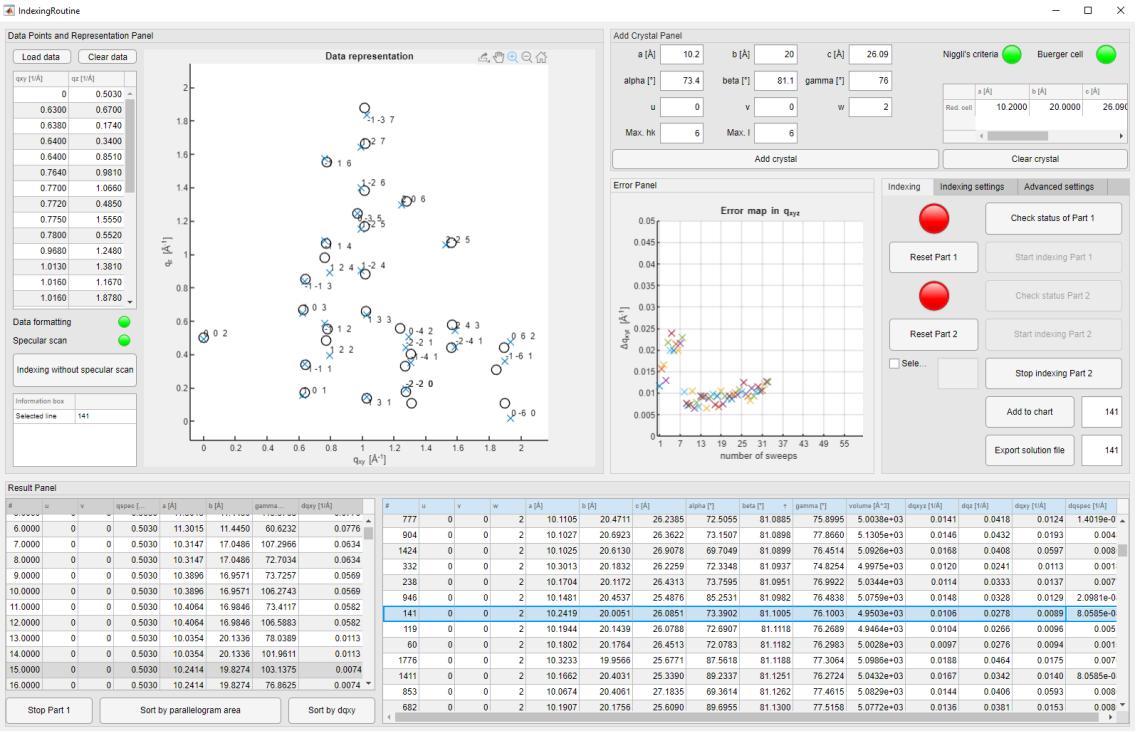


Figure 2.3: Screenshot of the main window of the Indexing Routine upon indexing of the provided data. The emphasized solution from the *Result Panel* is used to generate a simulated diffraction pattern. The pattern is added to the *Data Points and Representation Panel* for graphical comparison.

The best graphical agreement is recorded for a unit cell solution with the following parameters $u = 0$, $v = 0$, $w = 2$, $a = 10.24$ Å, $b = 20.01$ Å, $c = 26.09$ Å, $\alpha = 73.4°$, $\beta = 81.1°$ and $\gamma = 76.1°$. The volume of the reduced cell is $V = 4950.3$ Å^3 and the mean deviation in q_{xyz} is 0.01 Å^{-1}. A screenshot with the Indexing Routine's main window containing the as discussed solution is printed Figure 2.3 above. The overall time from loading the diffraction data down to a unit cell solution amounts to approximately one hour in this particular case. It has to be emphasized that this solution is not the only possible solution. No other solution than $(uvw) = (002)$ is observed for the contact plane with these settings. Sets with lattice constants similar to the one obtained appear frequently with only small differences in the numerical values. These differences, however, could lead to differently assigned Laue indices what makes the diffraction patterns overall different. Indexing and the task of searching a unit cell solution with this program is therefore an iterative exercise where the obtained possible solutions have to be verified eventually by the user. The routine does provide different numerically fitting solutions. A unique or indisputable solution for the provided GIXD data, however, can not be derived by application of this algorithm. In general, it is advised to compare the diffraction data

with the reciprocal space map in GIDVis, as demonstrated for one sample from the tutorial. For the sample of a Naproxen thin film and the two following samples, however, no data are available for a comprehensive comparison with the associated reciprocal space maps.

With this fourth example, this chapter is complete. Application of the Indexing Routine and the accompanying memory usage did not lead to any overflow or computational restrictions whatsoever. If the program parameter are kept in the ranges as demonstrated, the program should not face any programmatic problems. Nevertheless, no claim is made to completeness. Neither should the impression arise that the program has no limitations or bottlenecks, nor that the obtained solutions are unambiguous.

References

[1] B. Schrode et al. "*GIDVis*: a comprehensive software tool for geometry-independent grazing-incidence X-ray diffraction data analysis and pole-figure calculations". In: *Journal of Applied Crystallography* 52.3 (2019), pp. 683–689.

[2] J. Simbrunner et al. "Indexing of grazing-incidence X-ray diffraction patterns: the case of fibre-textured thin films". In: *Acta Crystallographica Section A* 74.4 (2018), pp. 373–387.

[3] J. Simbrunner et al. "Indexing grazing-incidence X-ray diffraction patterns of thin films: lattices of higher symmetry". In: *Journal of Applied Crystallography* 52.2 (2019), pp. 428–439.

[4] J. Simbrunner et al. "An efficient method for indexing grazing-incidence X-ray diffraction data of epitaxially grown thin films". In: *Acta Crystallographica Section A* 76.3 (2020), pp. 345–357.

[5] Stefan Kowarik et al. "Dewetting of an Organic Semiconductor Thin Film Observed in Real-time". In: *Advanced Engineering Materials* 11.4 (2009), pp. 291–294. DOI: 10.1002/adem.200800289.

Appendices

Appendix A

Diffraction data for indexing

A.1 fIna-04

Table A.1: Used peak positions for indexing fIna-04

q_{xy} [Å^{-1}]	q_z [Å^{-1}]
0.00000000	0.71115071
0.00000000	1.42145995
0.43687415	1.77769534
0.44482484	1.06574358
0.54148678	2.13065968
0.54226213	0.71026371
0.54250230	1.42057630
0.70075570	1.42372840
0.70843278	0.71058747
0.86269051	2.86649740
0.88529146	1.77676139
0.89009197	2.13067395
0.89216683	0.70901601
0.89374256	1.06110155
0.89555585	1.41849222
1.08173690	1.42493430
1.29845207	0.71006017
1.40855533	1.06057100
1.41827354	0.71226618
1.42377737	1.75819980
1.47928481	2.46386187
1.51842944	1.42000181
1.52069691	0.70863122
1.63072165	0.70999558
1.77333930	0.72669570
1.77896341	1.42525922
2.16200556	0.71334928
2.28577544	0.71354625

A.2 Diindenoperylene

Table A.2: Used peak positions for indexing DIP

q_{xy} [Å^{-1}]	q_z [Å^{-1}]
0.000	1.776
0.368	0.091
0.396	1.160
0.428	1.069
0.689	0.984
0.734	0.182
0.806	1.232
1.140	0.171
1.223	0.083
1.529	0.790
1.625	0.697
2.274	0.953

A.3 Pentacenequinone

Table A.3: Used peak positions for indexing PQ

q_{xy} [Å^{-1}]	q_z [Å^{-1}]
0.0000	1.9460
0.4520	1.3982
0.4550	0.5461
0.7740	1.9962
0.7810	0.0559
0.8850	1.3422
0.9090	0.8521
0.9090	1.0892
0.9120	1.4512
0.9140	0.4901
1.1750	1.1432
1.1790	0.7981
1.2160	1.0342
1.2170	0.9081
1.3640	1.6322
1.3670	0.3090
1.5460	1.6862
1.5580	0.1096
1.5980	0.3620
1.6050	1.2872
1.6080	0.6521
1.6340	1.5062
1.6380	0.4361
1.7750	1.1972
1.7760	0.7441
1.8230	0.2359
1.9560	0.2920
2.0100	0.1818
2.1080	0.4201

A.4 Naproxen

Table A.4: Used peak positions for indexing Naproxen

q_{xy} [Å^{-1}]	q_z [Å^{-1}]
0.000	0.503
0.630	0.670
0.638	0.174
0.640	0.340
0.640	0.851
0.764	0.981
0.770	1.066
0.772	0.485
0.775	1.555
0.780	0.552
0.968	1.248
1.013	1.381
1.016	1.878
1.018	0.881
1.018	1.667
1.024	0.659
1.026	0.137
1.240	0.559
1.269	0.333
1.275	0.177
1.280	1.318
1.307	0.403
1.312	0.107
1.563	1.070
1.563	0.444
1.568	0.577
1.846	0.311
1.894	0.444
1.899	0.107