Search and Identify with SIeve/SIeve+

A general utility program designed to search and identify materials from their characteristic powder diffraction patterns using the PDF-4 family of database products.

International Centre for Diffraction Data

ICDD®
INTERNATIONAL CENTRE FOR DIFFRACTION DATA

Dear Customers,

Welcome to the **SIeve**/**SIeve+** Technical Bulletin. This bulletin outlines the basic features of **SIeve**/**SIeve+** and can also be used as a quick user's guide. **SIeve**/**SIeve+** works seamlessly with PDF-2/PDF-4 databases, respectively. Both the databases and software are improved on an annual basis. The capabilities described here relate to Release 2013 products.

The Powder Diffraction File has been published since 1941. The file itself was part of a system developed for phase identification and chemical analysis and was intimately integrated with a classification and indexing system (Hanawalt et al., 1936, 1938). The indexing and phase identification system was called the Hanawalt method, named after one of the original inventors of the system, and has been in continuous use since 1941. Up until the 1940's, many scientists had collected reference data into databases. It was the development of the system that led to accurate phase identification.

The programs, **SIeve** and **SIeve+**, are **S**earch and **I**ndexing programs designed for use with PDF-2 and PDF-4 databases, respectively. The original concept of having a reference database with indexed and classified data intimately linked for phase identification and chemical analysis is embodied within **SIeve** and **SIeve+**. Both programs are based on a common platform, but each has an intimate link with the data and program interfaces with the unique contents of each database, **SIeve** for PDF-2, and **SIeve+** for PDF-4 databases. PDF-4 databases contain atomic coordinates and other crystallographic data and all data are expressed as digital patterns, so **SIeve+** has additional graphics and analysis features based on this content.

SIeve/**SIeve+** should not be confused with the original 1941 product. These programs work with JAVA® point and click interfaces, with standardized and edited data housed in a Sybase relational database. As shown in this technical bulletin, the programs have graphic interfaces and a wide variety of diagnostic tools that can be used for phase identification and semi-quantitative analysis. The original Hanawalt algorithm has been modified and enhanced to increase identification efficiency. This algorithm has been supplemented by additional phase identification algorithms (Long 8, Fink, Integral Index, Electron Diffraction, Similarity Index) that provide new capabilities for the identification of specific material classes or trace phase identification. The intimate link with the database means that the user can search on any specific custom data group that can be created through the numerous data mining searches and filters embodied in PDF-2 and PDF-4+.

SIeve and **SIeve+** are designed as a teaching tool, where the user has control over all the data processing and analysis steps. This enables greater control and understanding of the phase identification process, but at the loss of batch automation and simplicity. We hope the guide helps our users take advantage of the technical capabilities embedded within the database and identification software.

In the last few years, **SIeve** and **SIeve+** have undergone significant development. This development is based on user input that we received from emails, clinics, workshops, and our membership. If you have any comments or suggestions, please do not hesitate to call us.

Our goal is to help you solve your material problem!

Search and **I**dentify with **SIeve/SIeve+**

SIeve+ is a general utility program designed to search and identify materials from their characteristic powder diffraction patterns using the PDF-4 family of database products. **SIeve** is an analogous program used with PDF-2. The program was originally developed and released in 2003 (Faber et. al, 2004) to be an electronic replacement for the historic paper search and index manuals. Unlike its electronic predecessor, PCPDFWIN, this program was designed to interface with a modern relational database. Specifically, the program can use the extensive data mining interfaces, searches, and sorts available in PDF database products. A major revision in 2006 enabled JAVA® point and click interfaces for ease of use. As shown in the cited references, new features are added to **SIeve+** for PDF-4+ and **SIeve** for PDF-2 on an annual basis. Examples in this user guide were developed using the capabilities present in PDF-4+ 2013.

SIeve and **SIeve+** have undergone continuous improvement based on customer feedback. To enhance **SIeve+**'s capability, algorithms are updated as new data and features are added to the annual PDF-4 databases (Scardi et al., 2006; Faber and Blanton 2008; Kaduk and Reid 2011; Reid et al., 2011).

If you have any questions or suggestions about improving **SIeve+**, please contact ICDD at **info@icdd.com**.

Features

The strongest feature of the **SIeve+** program is its ability to intimately link to PDF-4+ filtering and data mining capabilities (Kabekkodu et al., 2002; Fawcett et al., 2011). This filtering improves the accuracy and precision of the identification process.

SIeve/SIeve+ uses **ALL PDF**:

- Subfiles
- Elemental filters
- Nomenclature filters
- Structural classifications
- Quality mark filters
- Physical property filters

For Release 2013 databases, **SIeve+** can filter with 57 searches. With all PDF databases, any search or combination of searches, can be used by **SIeve** or **SIeve+** as a user defined searchable database. OEM and independent distributor software typically interfaces to ICDD's subfile and elemental data in the database, but **SIeve+** interfaces with ALL the available information. In this sense, PDF-4 databases using **SIeve+** are far more powerful than their predecessors in helping you make an accurate identification.

Search Algorithms

Most of today's commercial software systems are founded upon the excellent and successful algorithms developed by Nusinovici and coworkers (Caussin et al., 1987; Nusinovici et al., 1988, 1992, 1993). Proprietary improvements have been added by various software developers to customize these algorithms to specific instruments and measurement operations. These algorithms are very powerful with demonstrated success in international round robin testing (Le Mains et al., 2003). The algorithms provide excellent results for highly crystalline materials and mixtures of highly crystalline materials. In commercial analysis systems, many of the proprietary software developments have optimized these algorithms for laboratory use under "normal" conditions. This has enabled automated fast batch processing and high throughput in laboratories around the globe.

Algorithm	What it does	When to use it
Hanawalt	Matches the **8 most intense (strongest) d-spacings** in the reference pattern to the experimental data. This is an extension of the historic 1930's algorithm that used the 3 most intense d-spacings.	Well proven general all purpose algorithm.
Fink	Matches the **8 longest and strongest** d-spacings. Sorts the most intense peaks by largest d-spacings. Originally developed in 1950's for organic materials (large d-spacings) and electron diffraction analysis. Performs best with large unit cells - organics, clays, zeolites.	Best for d-spacings > 5Å . It takes advantage of known d-spacing distributions.
Long 8	Matches the **8 longest** d-spacings independent of intensity. Developed in the 2004 Release to coincide with improved detector capabilities for weak intensities. Performs best with large unit cells.	Best with high quality data where you can identify weak peaks. Works poorly with low intensity data (below 1000 counts).
Electron Diffraction	Designed to work with 2D data from selected area electron diffraction (SAED) input. Assumes that the data are from different zones so it uses all data in the reference pattern and ignores intensity. Matches reference d-spacings in the user specified area. This was developed in 2012.	Specific to electron diffraction, but could also be used with any 2d data. This algorithm looks at all reference peaks in the defined regions. It is slower because it is computationally extensive.

*Table 1. Algorithms used in **SIeve+**. These are user selectable with Hanawalt as the default choice.*

As shown in Table 1, **SIeve+** allows users to customize identification algorithms on the type of data, instrument conditions, and various options. Most of the algorithms were developed and optimized for a particular type of analysis and, the selection options provide the user with a wide breadth of analysis capability. The various search algorithms, and their uses, are explained in Table 1, and the algorithm options are described in Table 2. The algorithm options shown in Table 2 follow some of the critical improvements previously described in the papers by Nusinovici et al., and provide enhanced capability to some of the historically proven Fink and Hanawalt algorithms (Hanawalt et al., 1936, 1938; Hanawalt, 1986).

Since the fundamental algorithms used by **SIeve+** are different than those used by most commercial software programs, they can be strongly complimentary. In the last few years, several publications (Fawcett et al. 2012; Kaduk, 2009; Gilmore et al., 2014) have shown the complimentary nature and demonstrate the strength of **SIeve+** in the analysis of minor and trace phases where the full filtering capability of the PDF-4 database family can be applied.

Algorithm Options	Toggle	What it does	When to use it	
Weight d-Spacings	OFF	8 lines equally weigthed, each contributes 0-1000 for a Max GOM score of 8000.	Reference patterns with many d-spacings in multiphase experiments.	Complex low symmetry materials.
	ON	Stronger lines contribute proportionally more to the GOM, but with the same maximum sum of 8000.	Reference patterns with few d-spacings in multiphase experiments.	Simple High symmetry materials.
2nd Pass Filter	OFF	No Penalty for a missing strong reference d-spacing.	Oriented specimens, samples with needles, platelets, etc.	
	ON	GOM penalty for a missing strong d-spacing in the experiment from the reference. Reduces candidates above GOM limits.	Well ground powder samples where orientation should not be a factor.	

Table 2. Options available that can be applied to the algorithms in Table 1. These affect the GOM scores for the selection of the top candidates in the identification process.

Many of the algorithms, and particularly the Hanawalt and Fink algorithms, are named honoring pioneers in materials analysis by powder X-ray diffraction. However, as applied by **SIeve+**, these algorithms are significantly advanced over those used in earlier decades.
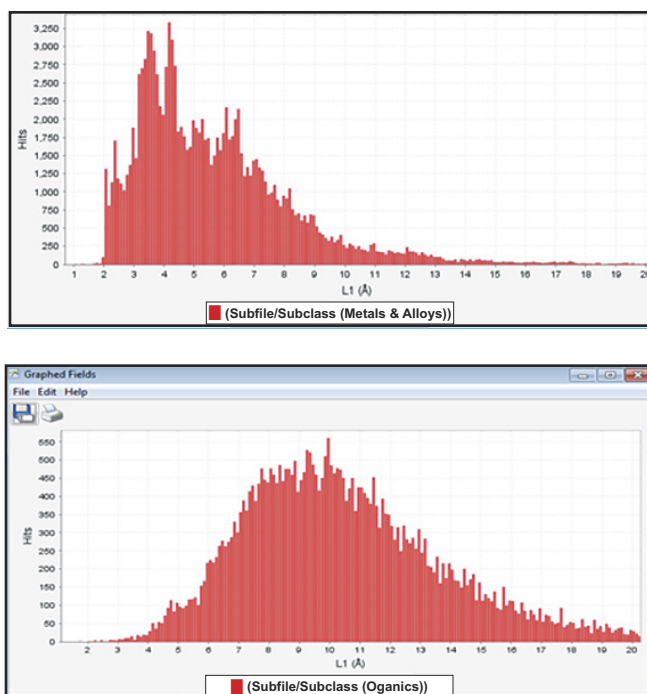

Figure 1. Long d-spacing (lowest angle), L1, distributions in metals and alloys (top) and organics (bottom).

The Fink algorithm searches the longest lines (L) in a diffraction pattern, those typically at very low angles. The longest d-spacings are usually indicative of a large unit cell. The d-spacing populations, as shown in the graphs above, can be significantly different for different classes of materials. A Fink search can be advantageous for materials with low d-spacing population and large unit cells with a long L (large) d-spacing(s). As originally developed, the Fink search was often a favorite of electron diffractionists because of the emphasis on low angle data and the limitations in two theta ranges in many electron diffraction systems.
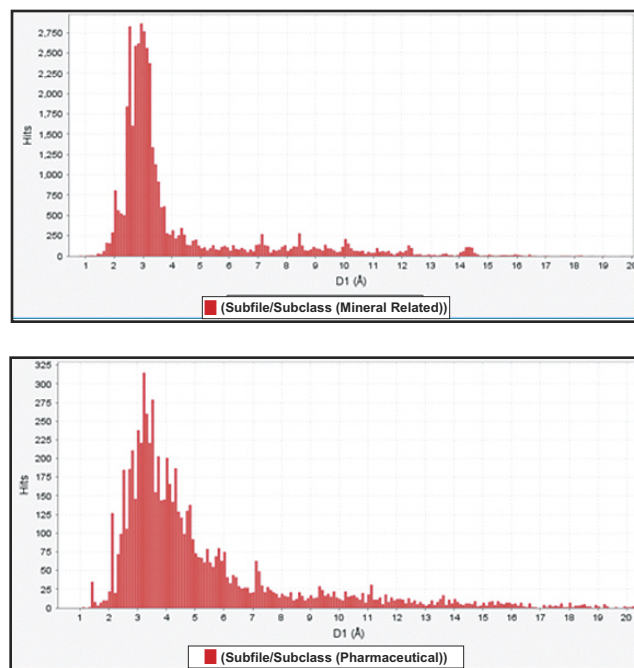

Figure 2. Strongest line (most intense), D1, in minerals and pharmaceuticals.

The Hanawalt algorithms search on the most intense d-spacings (D). The distributions of d-spacings for minerals and pharmaceuticals are shown above. Using the Hanawalt system, a mineral with a large d-spacing might be uniquely identified using only a couple of diffraction peaks due to the small statistical population at >5 Å.

These algorithms all use an index system first developed in 1936 by Hanawalt and coworkers.

The system cross references multiple d-spacings to quickly identify a material, and then uses the full pattern to confirm the analysis. As explained by Hanawalt, (Hanawalt et al., 1938, 1976, 1986) the three permuted line index was initially developed from a trial and error process using X number of permuted lines. This process concluded that three permutations was the optimum number needed to accurately analyze a phase in the shortest period of time. Finding the optimum number was very important from the 1930's to the 1980's.

Phase identification was a time consuming manual process. Since the 1930's, the precision of d-spacing has improved from parts per thousand to parts per million, and the number of reference patterns has grown from one thousand to over 800,000. Faber et al., 2004 recognized the problems with large databases with tens of millions of d-spacings and expanded the indexing system to permutations of eight d-spacings. The large permuted indexes which doomed the printed Fink Index in the 1960's, are relatively simple to implement in modern PC's with today's data storage and CPU capabilities. In subsequent years, ICDD scientists developed algorithms that use much larger indexes, such as the electron diffraction algorithm or full digital patterns such as the Similarity Index (Normalized R-index) described in Table 3 (Faber and Blanton, 2008). Due to their computational intensity, these algorithms, while very powerful, are visibly slower in CPU time.

Preference Options	What it does	Range	When to use it
Pattern GOM	Compares the reference data to the experiment, opposite of the GOM.		Very complex experimental patterns (>50 d-spacings) where strong overlapping peaks are expected. Weak and minor phase analysis.
Lowest allowable GOM	User defined GOM score needed to display.	>4000	High quality data with strong signal to noise (5-8 line match).
	Based on the quality of the d-spacing match.	2000 (Default)	Moderate quality data (2-6 line match).
		0-2000	Minor and trace phase identification (0-2 line match) best used during *residual pattern analysis* for minor phases.
Similarity Index, the Normalized R-Index	Uses digital patterns to compare all references to an experimental digital pattern. The algorithm does not depend on specimen crystallinity.	0 to 1	*Best for non-crystalline materials* (polymers, clays, some drugs) In this analysis a perfect point-by-point match would have a score of zero, scores under 0.5 are considered good.
	Does a point-by-point analysis comparing selected references to a user defined experimental data set.	>1	Inappropriate for complex multiphase specimens where the data would not be expected to match a single reference. Scores above 1 indicate that much of the experimental intensity was not found in the reference.

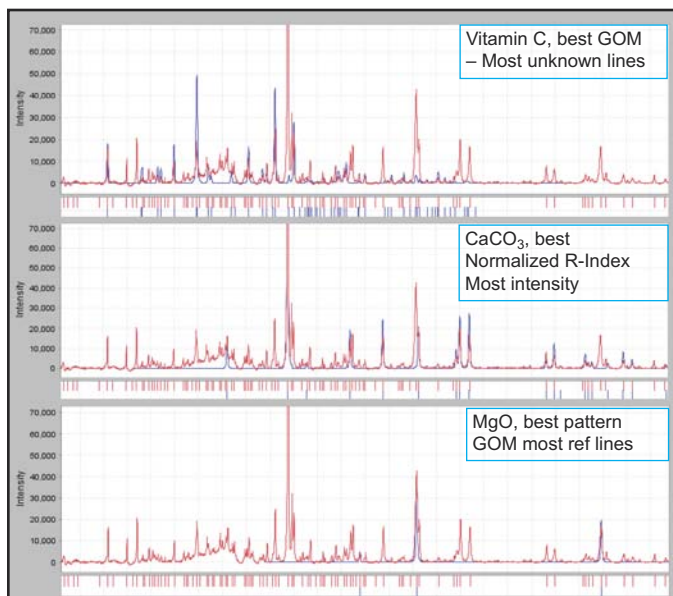Table 3. Options provided in "Preferences" for material identification.

Figure 3. Visual demonstration on how various algorithms would identify materials in a multiphase vitamin pill. The different algorithms identify different materials based on the mathematics of the algorithm.

Figure 3 shows the result using these different methods on the powder diffraction data obtained from a multiphase vitamin pill. Specifically, the illustration shows the initial best match found in each case. For a Hanawalt search (Faber et al., 2004), the best candidate, based on an unweighted Goodness of Match (GOM), was vitamin C. While this was not the strongest phase, all eight reference lines were matched by the experimental data resulting in a high GOM. The GOM is defined by the following equation, where the summation is performed for eight reference lines with a maximum GOM of 8000.

$$\text{GOM} = 1000 \cdot \sum \left[1 - |\delta d|/\text{SW}\right]^2,$$

where SW is defined as a delta two theta.

$$\text{SW} = \Delta(2\,\theta_{\text{SW}})$$

The normalized R-index (Faber and Blanton, 2008) identified $CaCO_3$. This phase is highly crystalline and the single phase that has the highest % of diffraction intensity in the experimental data.

$$S_r = \sum_{i=1}^{n} \left| \frac{I_i{}^{count}}{\sum_{j=1}^{n} I_i{}^{count}} - \frac{I_i{}^{calc}}{\sum_{j=1}^{n} I_i{}^{calc}} \right|$$

Since the Normalized R-Index shown above is a point-by-point analysis of diffraction intensity, $CaCO_3$ was the best match.

The highest score by the pattern GOM was with MgO. The reference pattern of MgO has only three peaks in the measurement range so the standard GOM score is very low compared to phases with eight or more peaks. However, the pattern GOM is very high since all three known peaks (out of three) were observed in the experimental data. The normalized R-index and pattern GOM provides the user with a capability not found in most commercial software systems. The specific equations used by these algorithms are described in the product help files.

Since the Normalized R-Index is based entirely on a point-by-point analysis between experimental data and reference scans, it has a singular advantage over all other algorithms. This algorithm can be used to analyze non-crystalline and amorphous materials where the diffraction pattern can exhibit both coherent and incoherent scatter. Thus, it becomes the method of choice for analyzing amorphous materials, such as polymers and nanomaterials. It also can be used with various nanomaterials having broad peak profiles when combined with reference nanomaterial data using the Scardi algorithm (Scardi et al., 2006). An example of a nanomaterial analysis is given in the case history on nanomaterials in the PDF-4/Organics technical bulletin.

Importing Data

Diffraction data are easily imported into **SIeve+** for PDF-4 products using a flexible file importer.
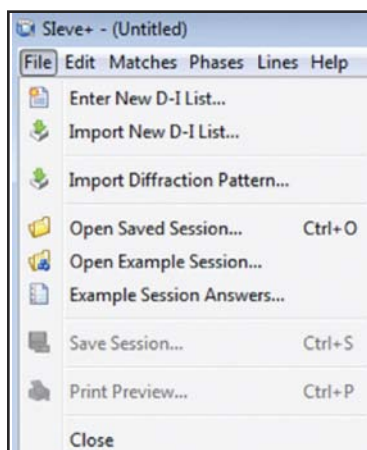

Figure 4. Data import selections include user data, saved data files and worked examples.

As shown in the figure above, the drop down menu under "File" presents many options for importing full diffraction patterns, d,I pair listings and previously saved data analysis sessions.

Under "Import Diffraction Pattern" the user can import diffraction data stored on their PC. The import program automatically recognizes several types of common data files from laboratory instruments and *.gsas files formatted for GSAS Rietveld refinement. More file types are added each year.

If the embedded importer does not recognize the file type, it will display the data and then wait for guidance from the user.
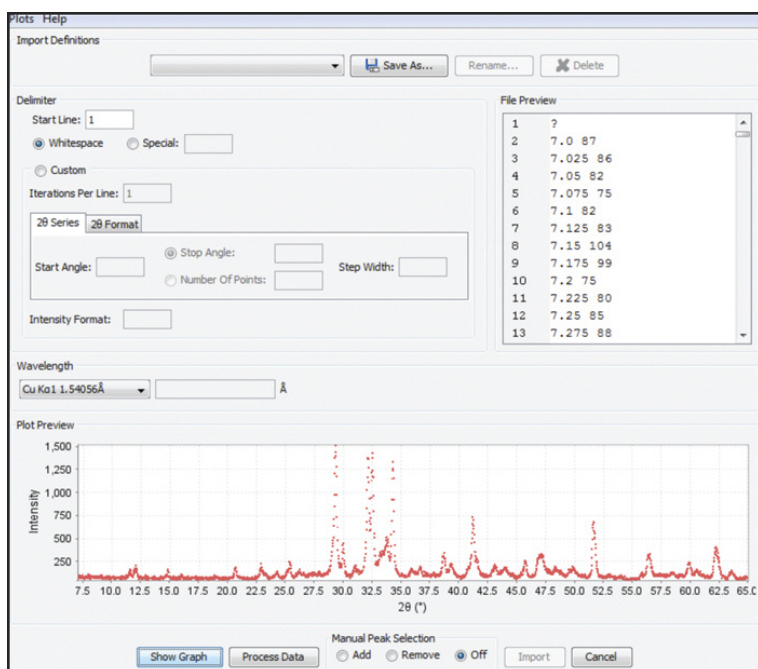

Figure 5. The data file import interface.

In Figure 5, a simple text file was imported, and the data (angle and intensity) are shown in the top right panel. The user must confirm the radiation used, specify the appropriate wavelength from a drop down box, and then the data are displayed. If this file had experimental conditions in a header, you can "teach" the importer to skip the header, describe the intensity format, and then save the format type. This process only has to occur once, then the stored format can be used to automatically import the data the next time.
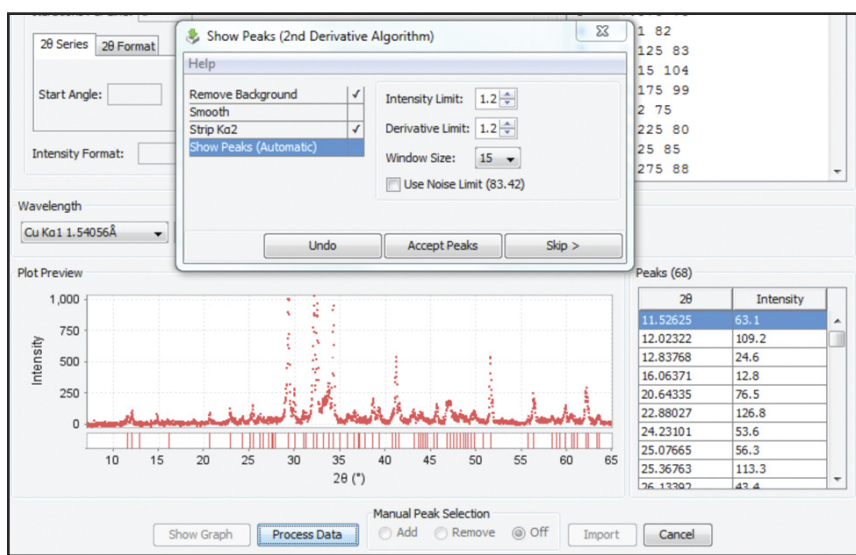
Data Processing

Figure 6. Data processing module.

Once the format is recognized, the user is asked to process the data. As shown in Figure 6 above, the data processing module uses a graphical interface where users have the option to remove background, smooth data, strip Kα2 intensity, and locate peaks. This produces the d,I pair list shown on the bottom right in Figure 6. Once the processing is finished, the data are imported for analysis. Each data processing step is graphically shown, but the user may skip steps if they have higher quality data, or adjust the algorithms to adjust for various experimental conditions. For example, if the user used an incident beam Ge monochromater, then the α2 stripping would be skipped.

The ICDD program developers deliberately kept each step with its own graphics interface (versus batch processed) so the user can view the quality of the processing at every stage and adjust algorithm variables as they desire. Industry standard algorithms are used with each processing step with "standard" default settings. Data processing is usually done with default parameters; however, the user can change these parameters depending on the type of diffraction pattern. It is important to pay attention to this as sometimes peak finding using default parameters may find false peaks as the second derivative method is sensitive to noise in the data. The graphical interface also provides features to add/remove peaks manually. Standard default

settings assume that the specimen has been ground into a fine powder and the data collected in a relatively short timeframe (30-60 minutes) on a Bragg-Brentano geometry diffractometer.

Process Step	Algorithm
Background Removal	Sonneveld-Visser
Smoothing Data	Savitzky-Golay
Strip Kα2	Modified Rachinger
Show Peaks	2nd Derivative

Table 4. Default data processing algorithms.

ICDD's customer surveys show that ICDD database users use a wide variety of non-standard optics and equipment. **SIeve** (PDF-2) and **SIeve+** (PDF-4) were specifically developed to handle a wide range of non-standard experiments. This includes radiation and wavelength varied experiments (synchrotron, neutron and electron diffraction data), as well as both low and high resolution X-ray diffraction data. With an interactive graphical interface, the user can determine the appropriate data processing settings for the experiment.

Analyzing Results – Set Up

Step-by-step analyses using **SIeve+** are given in the case histories of PDF-4+ Technical Bulletins. There are also case histories using **SIeve**, interfaced to PDF-2, shown in the PDF-2 Technical Bulletin. A main difference in the two programs is the enhanced ability of **SIeve+** to use digital pattern processing and display in every step of the analysis, allowing the user to study phase identification, as well as quantitative analysis, crystallinity and crystallite size. In PDF-2, all analyses use d,I listings. Additional examples are given in the case histories of the cited technical bulletins, all of which used **SIeve** or **SIeve+** in the analysis.

Database Case Histories

PDF-2—SIeve
 Multiphase Mineral
 Cement Analysis

PDF-4+—SIeve+
 Core Drilling Specimen
 Rietveld Refinement
  of a Production Catalyst
 Crystallinity and Crystallite Size
  in Cellulose

PDF-4/Organics—SIeve+
 Centrum Performance
 Formulation Analysis of Lipitor
 Nanomaterials

In general, imported data are processed to generate d,I listings that are then compared to PDF entry lines (Indexes) used by the various algorithms. Numerous Index tables are embedded in PDF-4 databases and the software accesses these tables during the identification process.

The exception to this procedure is the use of similarity indices, which does whole pattern matching where simulated profiles are compared to the experimental pattern. As applied in **SIeve+**, the d,I listing is initially applied to screen candidates, and then each prescreen candidate has a simulated pattern calculated that is compared to the experimental data. Similarity indices can also be calculated without using **SIeve+** through options in the PDF-4+ software. In this latter case, there is no d,I generation or prescreening step.

The user is presented with several options, which are all embodied in the "Preferences" icon in the header of the primary display screen.

On the "Simulated Profile" tab of the "Preferences" module, the user can select the radiation type, including neutron and electron wavelengths or energy. Preference choices include the optical geometry, a selection of various peak profiles, and a desired analysis range for any simulation. The "Preferences" module can be changed at any time during the analysis; there are default settings for all parameters. In general, the default settings assume you are doing a laboratory X-ray diffraction experiment, using Cu radiation, and a Bragg-Brentano optical configuration. The simulations will scale to the imported experimental data. If your experiment significantly alters from the "assumed" values, you can make the appropriate change to the preferences.

At the middle bottom of the form (Figure 7), one can see the button, "Set Synchrotron Defaults". This uses a set of default parameters that were experimentally determined at Advanced Photon Source of Argonne National Laboratories (Kaduk and Reid, 2011). Pushing this one button changes many default fields to values appropriate for the higher resolution of a synchrotron analysis. The user must input their specific wavelength.
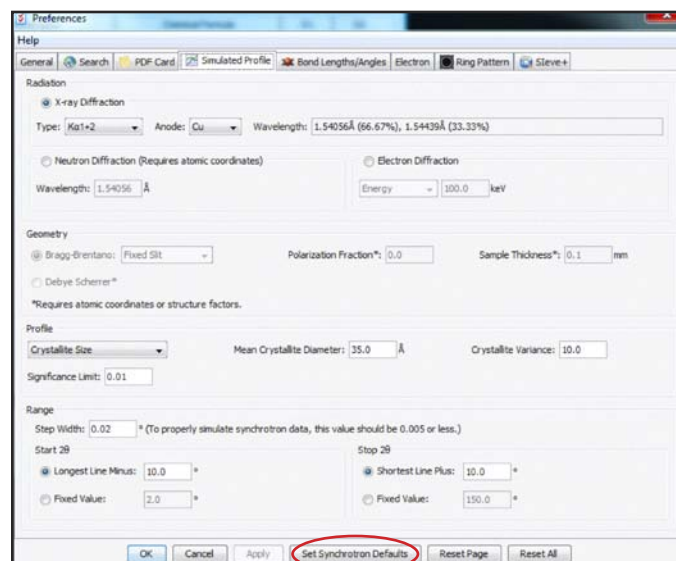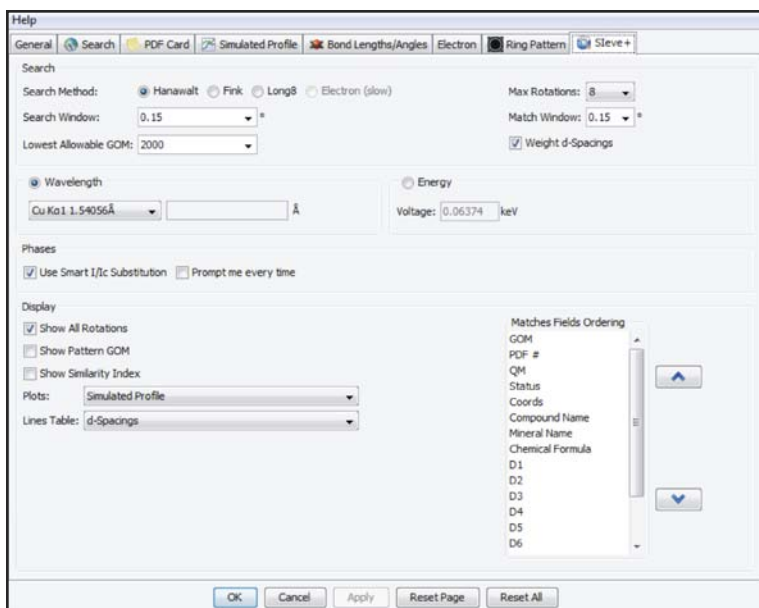

Figure 7. Preferences menu for PDF-4 database allowing the user to customize their analysis parameters.

Figure 8. Preferences menu in **Sleve+** allowing the user a selection of identification algorithms and search options.

Additionally, there is a separate "Preferences" menu specific for **SIeve+** as shown in the figure above. This is where the various search algorithms shown in Table 1 can be selected, as well as various options such as weighted d-spacing, GOM limitations, similarity index calculation, and pattern GOM calculation.

For the similarity index, it assumes that all reference data are compared with only the imported data for all calculations. In some commercial software systems, similarity indexes are applied in an N x N matrix in order to produce a cluster analysis of similar diffraction patterns. In **SIeve+** the similarity index compares a single experimental pattern to a series of references for the purpose of identification.

The bottom right selection box (Figure 8) allows the user to arrange various diagnostic fields in the display for matches and candidates.

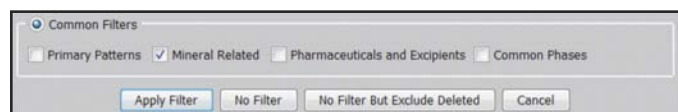## Analyzing Results — The Use of Filters

Figure 9. Common subfile selections in **Sleve** and **Sleve+**.

Upon data import, the above screen will be displayed. The user is given an option to use the full database (No Filter) or use several common filters.

The purpose of these common filters is to improve the accuracy of the analysis by restricting the search to materials of the appropriate quality and chemistry. Table 5 outlines the key attributes of each of the common filters. This is the value of an edited database where the materials are reviewed and classified. From the above table, one can see the effect of various filters on reducing the "searchable" entry populations.

Users typically know whether they are analyzing a soil sample or a pharmaceutical tablet, and this information can be used as a critical filter in your analysis. The filters also provide the user with the opportunity to exclude deleted patterns and alternate patterns through the selection of "primary" patterns. If the user has moderate or high quality experimental data, the selection of primary patterns is recommended. When there are multiple patterns available for a given phase, a higher quality pattern is chosen as "Primary" and the rest are flagged "Alternate". The reason for keeping low quality "Alternate" patterns in the database is low quality experimental data (often due to small sample size) often matches these lower quality references better than the more recent high quality references.

Filters	No of Entries	% PDF-4+	What it does	When to use it
No Filter	340,653	100	Includes all data.	You know nothing about the sample and have low quality experimental data.
No Filter, Exclude Deleted	325,603	95	Removes deleted data that have large known errors.	You know nothing about the sample.
Apply Filter Primary Patterns	259,985	76	"Best" example of a particular chemistry, removing all alternate data of the same chemistry.	Recommended default when you know nothing about the sample and have good quality data collected on a modern instrument.
Mineral Related	40,424	12	Classified Minerals	Comprehensive collection for minerals analysis.
Pharmacueticals and Excipients	7,266	2	Materials found in pharmacueitcal formulations.	When analyzing tablets.
Common Phases	20,641	6	Common materials as identified by ICDD members in the 1980's.	Includes common minerals, drugs and corrosion products. Modern advanced materials such as fuel cell components would not be included.

Table 5. Attributes of common filters used in **Sleve** and **Sleve+**.

These filter screens are offered at the beginning of an analysis, and also during an analysis as a convenience to the user. They provide fast access to commonly used filters through a point-and-click JAVA® interface.

The user also has the option to make a custom filter based on the data mining capability of the PDF databases. Any time a data mining search is performed during a user session, the search is saved in the history folder. The contents of the history folder are made available in **SIeve+** through the Matches Filter.
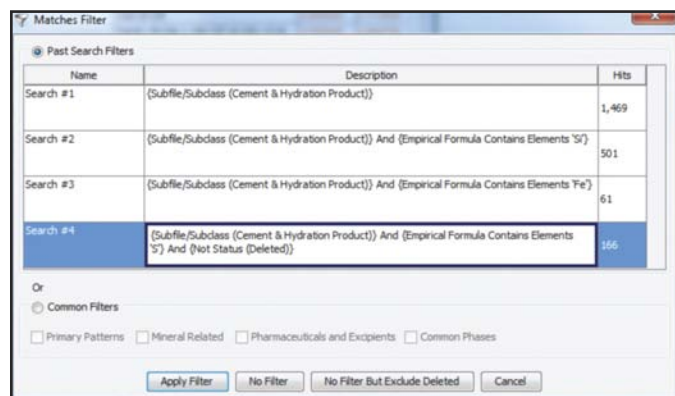

Figure 10. Matches Filters used with **Sleve** and **Sleve+**. The results of a Boolean-type user search during a single PDF database session will show up as selectable filter.

In Figure 10, the user was analyzing cement and had elemental data indicating bulk concentrations of calcium, silicon, iron and sulfur. Boolean searches were run in PDF-4+ using the "Cement and Hydration Product" subfile, and then subsets that contained Si, Fe and S. Each of these searches is shown above in the filters selection of **SIeve+**. The user simply highlights the selection to use it. The number of entries for each selection is shown on the right in Figure 10. One can see that this type of filtering can be highly effective in selecting targeted chemistry, which uses less than 0.5% of the database entries and identifying the correct phase in each case!

Top phase identified

Cement	Ca_3SiO_5
Cement & Si	Ca_3SiO_5
Cement & Fe	$Ca_2(Fe, Al)O_4$
Cement & S	$CaSO_4$

The minerals listed above are haturite, brown millerite and gypsum, which are common components in Portland cement. A full analysis of this nine phase cement sample is given in the case histories and several examples where subfiles and elemental searches were combined are provided in the references. These types of searches have also been described in recent publications as an effective method for trace phase analyses (Kaduk, 2009; Fawcett et al., 2011).

Subfile searches are especially effective. Each subfile has a specified definition for the chemistry of the subfile contents. Classification is performed by field experts and ICDD's editorial staff. Some subfiles, such as metals and alloys and minerals have been edited and improved by teams of field experts for over 60 years. A tabulation of each subfile in ICDD's major products, PDF-4+ and PDF-4/Organics is given in Table 6. Any of these subfiles can be used as a filter in the search/match process. PDF-4+ has comprehensive collections of ceramics, minerals, and metals and alloys, which are primarily based on inorganic chemistries. PDF-4/Organics has comprehensive collections for amino acids, carbohydrates, pharmaceuticals, steroids, and clathrates, which are primarily based on organic chemistries.

In the majority of case histories, a subfile filter was employed to improve accuracy in the phase identification process.

PDF4/Organics 2014	Subfile	PDF-4+ 2013
2,637	Alkaloids	371
10,361	Amino Acids, Peptides & Complexes	1,013
15,766	Bioactivity	1,462
36	Battery Materials	404
5,746	Carbohydrate	651
581	Cement Materials	1,469
387	Ceramics	12,833
2,228	Common Phases	20,641
404	Education	1,071
777	Explosives	1,230
4,886	Forensics	13,186
157	Intercalates	407
11,198	Inorganics	309,139
1	Ionic conductors	1,963
1,582	Merck	0
1,854	Metals and Alloys	126,042
7,019	Micro & Mesoporous	4,214
6,210	*Clathrate*	17
81	*Metal organic framework*	57
691	*Zeolites*	3,724
2,569	Mineral & Related Materials	40,424
25	*Gem*	77
1,632	*Minerals*	34,308
753	*Naturals*	20,922
879	*Synthetics*	13,386
2,240	Nucleosides & Nucleotides:	629
214	NBS	2,209
476,029	Organics	38,193
9,333	Pharmaceuticals	7,266
2,429	*Excipients*	2,476
861	Pigments	1,346
1,452	Polymers	1,277
5,029	Porphyrins, Corrins & Complexes	207
4,911	Steroids	297
292	Superconducting Materials	7,375
19	*Reaction Products*	297
7	*Related Materials*	515
2	*Conventional*	241
1	*High TC*	171
3,943	Terpenes	99
0	Thermoelectric Materials	941
5	Hydrogen storage materials:	1,778
17	Modulated Structures	286

Table 6. Subfile and subclass populations in two major PDF databases.

Analyzing Results – Phase Identification

The phase identification process begins automatically after the experimental data are imported and the filter selected. Top candidates are scored and displayed on the diagnostic display.
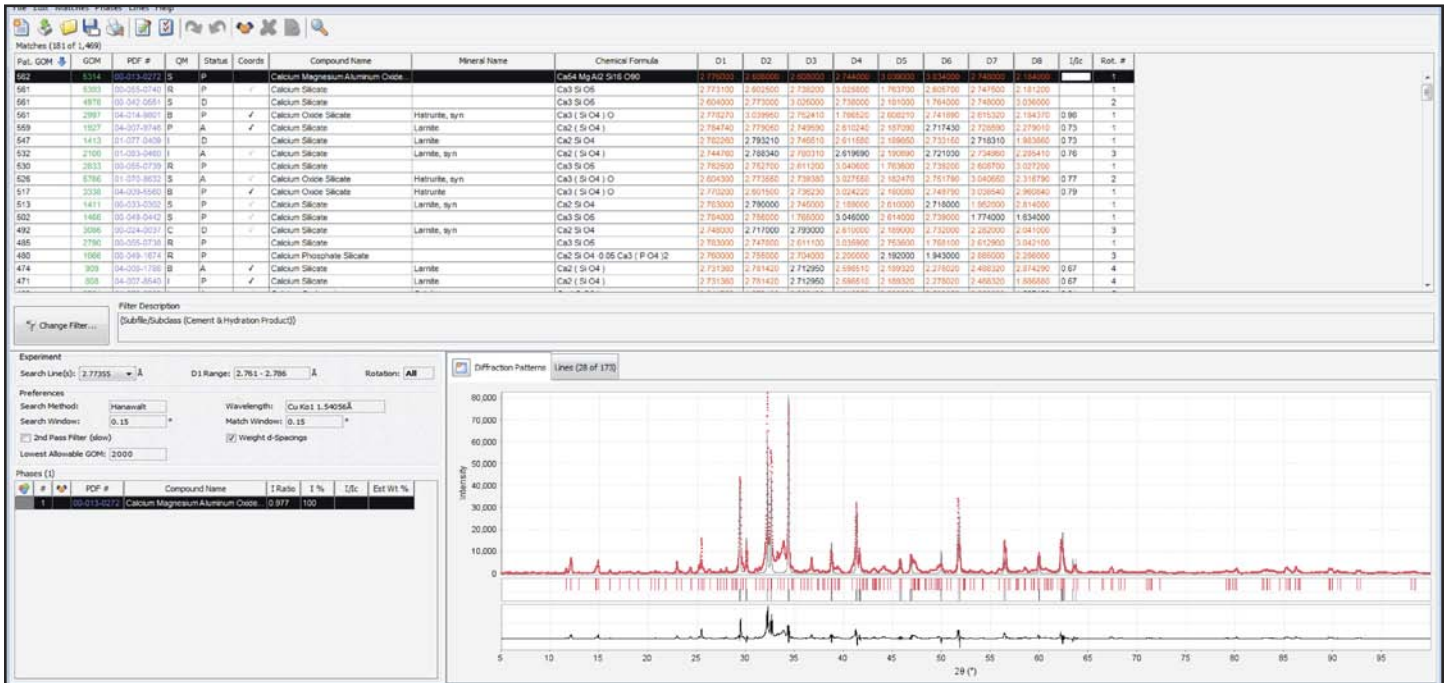
Figure 11. Diagnostic display for **SIeve+**.

The diagnostic display consists of three panels – top, bottom left and bottom right. The imported data are graphically shown on the bottom right panel, if using **SIeve+**.

Figure 12. Diagnostic display for **SIeve** showing a d,I list comparison in the bottom right panel. This is the default panel for **SIeve** and an optional panel for **SIeve+**.

Alternatively, the imported data can be shown as a d,I pair listing by clicking on the nested tab. The display, Figure 12, now shows the imported d,I listing on the bottom right panel.

File Edit Matches Phases Lines Help

Matches (132 of 1,469)

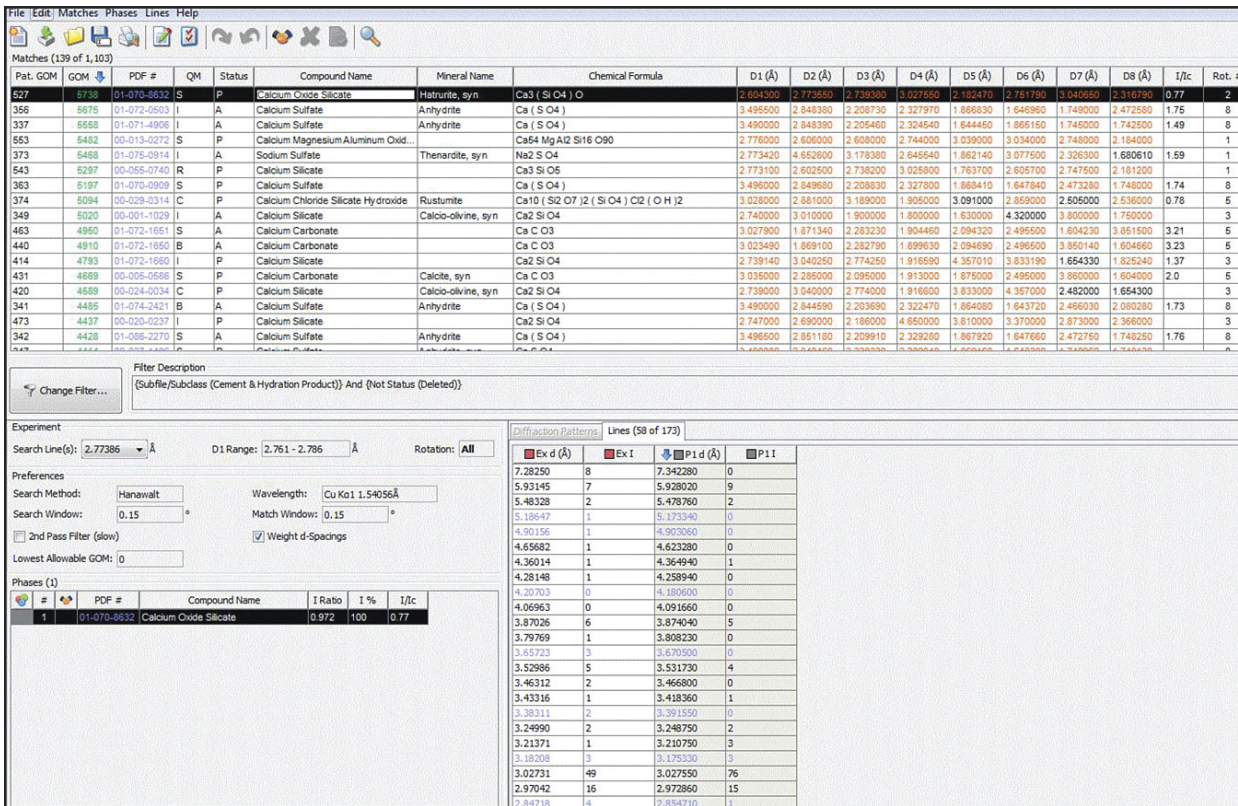| GOM	PDF #	QM	Status	Coords	Compound Name	Mineral Name	Chemical Formula	D1	D2	D3	D4	D5	D6	D7	D8	I/Ic	Rot. #
5766	01-070-8632	S	A	✓	Calcium Oxide Silicate	Hatrurite, syn	Ca3 (Si O4) O	2.604300	2.773550	2.739380	3.027550	2.182470	2.751790	3.040650	2.316790	0.77	2
5752	01-071-4906	I	A	✓	Calcium Sulfate	Anhydrite	Ca (S O4)	3.490000	2.848390	2.205460	2.324540	1.644450	1.865150	1.745000	1.742500	1.49	8
5745	04-008-0191	B	A	✓	Calcium Sulfate	anhydrite	Ca (S O4)	3.495500	2.848380	2.208730	2.327970	1.866830	1.646960	1.749000	2.472580	1.75	8
5624	01-075-0914	I	A		Sodium Sulfate	Thenardite, syn	Na2 S O4	2.773420	4.652600	3.178380	2.645540	1.862140	3.077560	2.326300	1.680610	1.59	1
5396	00-029-0314	C	P		Calcium Chloride Silicate Hydroxide	Rustumite	Ca10 (Si2 O7)2 (Si O4) Cl2 (O H...	3.026000	2.881000	3.189000	1.905000	3.091000	2.859000	2.505000	2.536000	0.78	5
5393	00-065-0740	R	P		Calcium Silicate		Ca3 Si O5	2.773100	2.602500	2.738200	3.025800	1.763700	2.605700	2.747500	2.181200		1
5314	00-013-0272	S	P		Calcium Magnesium Aluminum Oxide...		Ca54 Mg Al2 Si16 O90	2.776000	2.606000	2.608000	2.744000	3.039000	3.034000	2.748000	2.184000		1
4986	00-006-0226	O	D		Calcium Silicate	Anhydrite, syn	Ca S O4	3.498000	2.849000	2.328000	2.206000	1.869000	1.648000	1.749000	2.086000		8
4978	00-042-0651	S	D		Calcium Silicate		Ca3 Si O5	2.604000	2.773000	3.025000	2.738000	2.181000	1.764000	2.748000	3.036000		2
4957	01-080-0042	H	A	✓	Magnesium Carbonate	Magnesite, syn	Mg (C O3)	2.739590	2.106400	1.699870	1.695650	2.492830	1.943490	1.341760	1.490680	1.77	3
4821	00-002-0623		D		Calcium Carbonate	Calcite	Ca C O3	3.030000	1.920000	1.870000	2.280000	2.100000	1.600000	1.520000	2.490000		5
4607	04-012-5254	I	P		Calcium Chloride Silicate Hydroxide	Rustumite	Ca10 (Si O4) (Si2 O7)2 Cl2 (O H...	3.028240	2.884750	1.905000	3.188520	3.091670	2.504570	1.751420	2.859120	0.61	5
4607	04-016-3635	B	P	✓	Calcium Chloride Silicate Hydroxide	rustumite	Ca10 (Si O4) (Si2 O7)2 Cl2 (O H...	3.028240	2.884750	1.905000	3.188520	3.091670	2.504570	1.751420	2.859120	0.59	5
4604	04-002-9082	P	A	✓	Calcium Carbonate	Calcite, syn	Ca (C O3)	3.023060	1.867520	2.274710	1.904910	2.085190	2.483700	1.597000	3.838150	3.25	5
4596	01-074-1782	B	D		Calcium Sulfate		Ca S O4	3.490000	2.844590	2.203690	2.322470	1.864080	1.643720	2.466030	1.742500	1.73	8
4623	00-060-0311		P		Samarium Manganese Oxide		Sm0.95 Mn O2.925	2.737200	1.937900	3.870600	1.579200	2.244900	2.232400	1.588300	1.370900		3
4492	01-086-2270	S	A		Calcium Sulfate	Anhydrite	Ca (S O4)	3.496500	2.851180	2.209910	2.329280	1.867920	1.647660	2.472750	1.748250	1.76	8

Change Filter...

Filter Description
{Subfile/Subclass (Cement & Hydration Product)}

Figure 13. Top diagnostic panel.

The top panel, shown in Figure 13, contains a large amount of diagnostic information. The headers for each column are in the Table. The first column is the sorted GOM. The GOM is a measure of how close the d-spacings of the reference match those of the experiment for the eight most intense peaks in the reference pattern. Every d-spacing match, within the error limits set by the user, is colored in red. These matches are shown in the D1 to D8 columns on the right of the panel (Figure 13). In this case, for most of the top candidate phases, the selections are primarily red and the GOM values are high.

The program automatically selects the candidate with the highest GOM value. As seen in Figure 11, this candidate is highlighted in the top panel, highlighted in the bottom left panel, and shown as a digital pattern superimposed with the experimental data in the bottom right panel. In the case of d,I pairs, the d,I listing of the top candidate phase is shown in the bottom right panel of Figure 12.

In addition to the GOM, the program provides a PDF identifier code, quality mark (QM), status, and an atomic coordinates check if the reference contains atomic coordinates. Compound name, mineral name, chemical formula, and I/I_c are also displayed. The user has the ability to change the display headings by using the preferences module shown in Figure 8. Pattern GOM and normalized R-index values can also be selected.

For the top candidate phase, we matched Ca_3SiO_5, haturite. It has an "S" quality mark and an "A" status. "S" stands for star quality and designates that the chemistry, crystal structure, unit cell and physical properties for this material are all self-consistent. "A" stands for alternate pattern. The PDF number 01-070-8632 indicates that this material was from a single crystal database source, 01-ICSD, and that the entry contains atomic coordinates. Atomic coordinate data are important if the user wants to do a subsequent Rietveld analysis to quantitate the phases. The graphical fit of this reference (grey) matches very well with the experimental data (red), as shown in the bottom right plot panel(s) in Figures 11 and 14. One

can accept this phase using the 'Shake Hands' icon on the toolbar, or the user can choose a different candidate further down the list to accept. It is possible, using a 'Ctrl-click,' to choose multiple phases to accept at one time. Once the phase(s) is accepted, another search will begin using the experimental peaks that were not matched with the initial identified phase(s).
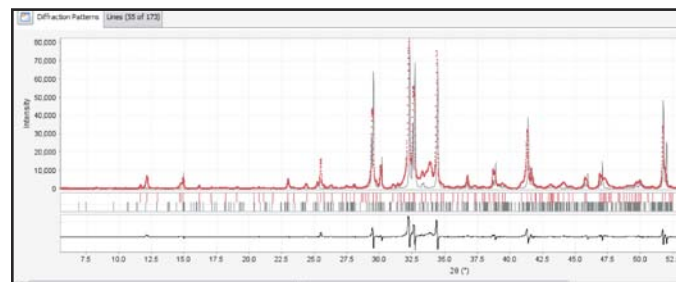

Figure 14. Diffraction pattern plot panel showing the experimental data (red), reference pattern simulation (grey) and difference plot (black).

The d-spacing matches for this first candidate entry are also quite close as shown in the full pattern comparison list in Figure 15. In fact, the full listing shows that 55 d-spacings from the reference are found in the experimental data. One can see by the listing that the experimental and observed d-spacing agree to within a part per thousand for the most intense peaks. This agreement is very good especially considering that no internal standards were used to calibrate the peak locations. (Fawcett et al., 2004). The peak intensities do not agree as well, except in the overall general magnitude of weak and strong peaks.

This example is "classical" in the sense that it shows excellent agreement in d-spacing measurements with less precise agreement in the intensity measurement. The d-spacing measurement can be determined to high precision, and if an internal standard is used, they can be determined to high accuracy. This has been shown in numerous publications on accuracy and precision in XRD analyses and in several international round robin studies.

Ex d	Ex I	P1 d	P1 I
4.27589	1	4.258940	0
3.87016	6	3.874040	5
3.79696	1	3.808230	0
3.52949	5	3.531730	4
3.46290	3	3.466800	0
3.38365	2	3.391550	0
3.28279	1	3.285050	1
3.24842	2	3.248750	2
3.21291	1	3.210750	3
3.18250	3	3.175330	3
3.08377	1	3.082180	2
3.02721	50	3.027550	76
2.97081	16	2.972860	15
2.84735	4	2.854710	1
2.82741	3	2.826420	2
2.80723	8	2.798390	3
2.77355	100	2.773550	98
2.74884	68	2.751790	50
2.73967	63	2.739380	91
2.69281	12	2.689870	4
2.66942	11	2.667450	2
2.60757	93	2.604300	100

Figure 15. D-spacing and intensity (d,I) comparisons for the experiment (Ex d,I) and identified phase (P d,I). Both are auto scaled to a maximum intensity of 100.

There are numerous physical contributions to the intensity and large variations should be expected. The accuracy of the d-spacings and variability in the intensities are based on the fundamental diffraction physics and covered in many textbooks as well as ICDD's educational clinics.

The practical consequence is that the match and identification process is heavily weighted towards positional d-spacing accuracy in all known commercial analysis systems, and with every algorithm! Intensity information is included particularly with whole pattern matching methods, but it is a secondary contributor.

So our diagnostic information on matching haturite to our experimental data includes a high GOM, graphical confirmation, d-spacing matches on 55 d-spacings, and use of the highest quality reference (star). This material can be positively identified with confidence.

In this same example, other candidates on the list, with high GOM's, included CaO and $MgCO_3$. However, the graphic displays and d-spacing matches would show that the intensity distributions were not accurate. For example, the largest peak of CaO matches a normalized experimental intensity of two and some weaker d-spacings were not observed. Because the

GOM analysis is based solely on d-spacing matches it is always recommended that the user does either a visual graphics comparison or d,I listing comparison check for every match.

The previously described Nusinovici algorithm utilizes both positional and intensity information and penalizes candidate references that have high intensity peaks in areas where the experimental data does not have peaks. **SIeve** and **SIeve+** incorporates this logic in the "2nd pass filter" option that is a new feature.

The "2nd pass filter" examines large intensity d-spacings in the pattern of the reference, and if they are missing in the experimental data, the GOM is penalized by subtracting a "penalty" score, effectively demoting the candidate in the candidate list. When the "2nd pass filter" is deployed, the number of matching candidates is always reduced. In the above example from the analysis of a cement, 75 candidates in the cement and hydration product subfile had GOM's >2000 in the candidate list. Use of the second pass filter reduced this list to 27. This removed CaO and $MgCO_3$ from consideration and, in effect, strongly promoted $CaCO_3$ as a candidate phase. Fourteen (14) d-spacings from the $CaCO_3$ reference matched the experimental data and graphic fits. The d,I comparisons were very good providing confidence to the identification.

Because the "2nd pass filter" is an intensity supplement to the GOM scoring of d-spacings, it should be used in most cases. However, the "2nd pass filter" would not be appropriate in cases where the user suspects preferred orientation, which can produce large variations in intensity.
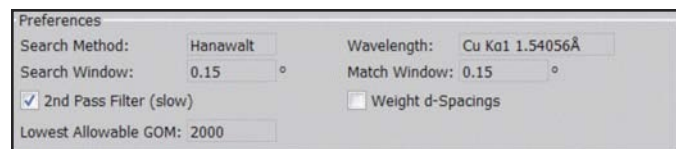

Figure 16. Preferences panel shown in the diagnostic display panels. This is an enlargement of the information shown on the bottom left of Figures 11 and 12.

Preference selections are shown in a preferences box that is part of the lower left panel of the diagnostic display screen. One can see that this "2nd pass filter" can be toggled on or off just by clicking the box (√). It is often instructional to compare identification results with the toggle on and off.

Another evaluation tool is also shown in the preferences selection, "weight d-spacings". This applies a weighting scheme to the reference data and weighs the most intense d-spacings in the eight d-spacing index. This is a "practical" addition and is useful in situations where the number of reference lines for a particular material is expected to be less than eight. This occurs for the following reasons:

1) The reference pattern is from a high symmetry material of a simple (1-2 element) composition.

2) The reference pattern is dominated by 1-2 high intensity d-spacings.

Examples of the former would be nearly all metals and their binary oxides. Examples of the latter would be quartz, Teflon, polyethylene, and calcite. In the latter case, if the material was present in the specimen at low concentration, then only the most intense peaks would be observed. This weighting increases the GOM for materials that have a few strong peaks. Unlike the "2nd pass filter", the weighted d-spacings option typically does not reduce the candidate list, but it will change some of the candidates and their GOM scores. This option can also be toggled ($\sqrt{}$) on and off in the display panel.

In addition to some of the search match algorithms discussed previously, **SIeve+** also allows varying the search match conditions like search window, match window, and rotating the lines.

Scoring Method	Order of phase ID
GOM	Calcite, monetite, vitamin C, ZnO
Weighted GOM	Calcite, monetite, KCl, vitamin C
2nd Pass GOM	Calcite, monetite, ZnO, KCl
2nd Pass, weighted	Calcite, monetite, KCl, ZnO

Table 7. Order of phase identification for various options using the same data set for a Centrum Performance vitamin pill.

The analysis of a Centrum Performance vitamin pill (Table 7) provides some insight into the scoring methods used for identification. The standard GOM easily finds vitamin C and monetite because they both have many d-spacings that match the experimental data, despite being weaker intensity phases in the experiment. The weighted GOM's promote KCl, which is the dominant (most intense) phase in the experimental pattern, but only has four d-spacings in the analysis range. Calcite is identified in every method since there are 10 d-spacings in the analysis range, and it is the second most intense in the experiment after KCl. MgO is also a component of the Centrum Performance Formulation, but it has only two d-spacings in the analysis range, which results in a low GOM even in weighted cases.

This complex formulation has >100 experimental d-spacings in a laboratory analysis. MgO is typically identified only after all other major phases have been accounted for, and the analyst is examining the remaining unidentified peaks in a residual analysis. (Fawcett et al., 2011)

Alternatively, if one knew that the vitamin pill contained magnesium (XRF, EDX, or simply from the list of ingredients), then a search of magnesium containing pharmaceuticals would find 77 candidates and the only candidate phase that would match the experimental data in **SIeve+** is MgO! In the case histories, the analysis of Centrum Performance usually results in the identification of 8-10 phases using laboratory methods and intelligent filter selection with **SIeve+**.

There are two additional tools that can be applied to any analysis. These are the pattern GOM and Similarity Index. The pattern GOM reverses the logic of the standard GOM so that instead of comparing the reference to the experimental pattern, the experimental pattern is compared to the reference. It uses the same eight d-spacing Indexes used in the standard GOM calculation. This can often be used to identify simple patterns in complex data sets where many of the reference peaks may be "hidden" or overlapped by the experimental data. In the case of the Centrum Performance, the pattern GOM easily identifies MgO and $CaCO_3$. In the case of the previously discussed cement specimen, the pattern GOM identifies calcium trisilicate (haturite) and calcium disilicate (larnite). While the haturite is the strongest intensity phase, the larnite having lower intensity, has many overlapping d-spacings, and is not as easily identified in the standard GOM. Our annual product testing would indicate that the pattern GOM is not as accurate as the standard GOM, but in the above two cases it would immediately direct the user to phases that are difficult to identify.

SIeve/SIeve+ and Commercial Software

The staff of the International Centre for Diffraction Data works with major software developers and equipment manufacturers around the world. We frequently develop database features at the request of software developers to enhance identification analyses with their proprietary software. Most commercial systems utilize the PDF's subfile and elemental filter systems that improve the accuracy of results. As previously mentioned, the algorithms used in most major commercial diffraction systems have been shown to be highly effective in analyzing crystalline mixtures (LeMains et al., 2002).

SIeve and **SIeve+** offer a variety of algorithms and options that allows users to optimize results for particular chemistries and both standard and non-standard diffraction systems. **SIeve/SIeve+** are fully integrated with the data mining capabilities of PDF databases providing extensive filtering to enhance the identification process. In many cases, this provides a unique capability to analyze the most difficult problems. In the last few years, new algorithms have been added to identify non crystalline materials and complex multi-phase specimens.
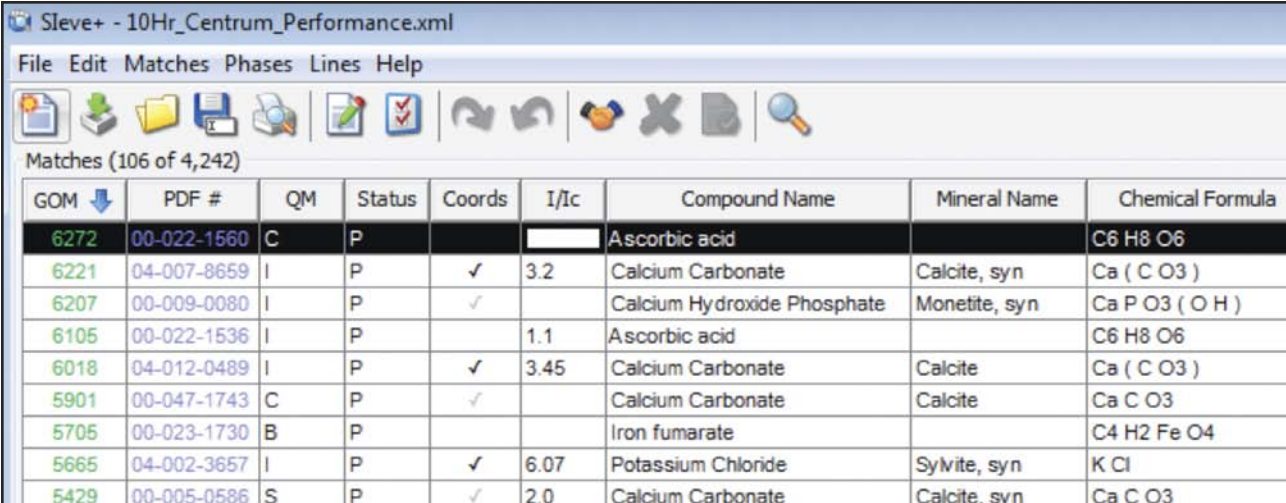
References

Caussin, P., Nusinovici, J., and Beard D. W. (**1987**), *"Using Digitized X-ray Powder Diffraction Scans as Input for a New PC-AT Search/Match Program,"* Adv. in X-ray Anal., 31, 423-430.

Faber, J., Weth, C.A., and Bridge, J. (**2004**), *"A Plug-in Program to Perform Hanawalt or Fink Search-Indexing Using Organics Entries in the ICDD PDF-4/Organics 2003 Database,"* Powder Diffraction, 19 (1), 26-30.

Faber, J. and Blanton, J. (**2008**), *"Full Pattern Comparisons of Experimental and Calculated Powder Patterns Using the Integral Index Method in PDF-4+,"* Adv. in X-ray Anal., 51, 183-189. These same two authors developed a normalization of the integral index (normalized R-index) that was published in PDF database products starting with Release 2011.

Fawcett, T. G., Kabekkodu, S. N., Faber, J., Needham, F., and McClune, F. (**2004**), *"Evaluating Experimental Methods and Techniques in X-ray Diffraction Using 280,000 Data-Sets in the Powder Diffraction File,"* Adv. in X-ray Anal., 47, 156-164.

Fawcett, T. G., Crowder, C. E., Kabekkodu, S. N., and Kaduk, J. A. (**2011**), *"Improved Material Identification Through Targeted Data Mining,"* Adv. in X-ray Anal., 54, 149-161.

Gilmore, C., Kaduk J., and Schenk H., and editors (**2014**), International Tables for Crystallography, Volume H Powder Diffraction, Published by the International Union for Crystallography.

Hanawalt, J. D. and Rinn, H. W. (**1936**), *"Identification of Crystalline Materials, Classification and Use of X-ray Diffraction Patterns,"* Ind. Eng. Chem. Anal., (8), 244. Reprinted (1986) Powder Diffraction, 1 (1), 2-7.

Hanawalt, J. D., Rinn, H. W. and Frevel, L. K. (**1938**), *"Chemical Analysis by X-ray Diffraction,"* Ind. Eng. Chem. Anal. (10), 457-512, Reprinted (1986) Powder Diffraction, 1 (2), 2-14.

Hanawalt, J. D. (**1976**), *"Phase Identification by X-ray Powder Diffraction. Evaluation of Various Techniques,"* Adv. in X-ray Anal., 20, 63-73.

Hanawalt, J. D. (**1986**), *"Manual Search/Match Method for Powder Diffraction in 1986,"* 1(1), 7-14.

Kabekkodu, S. N., Faber, J. and Fawcett, T. (**2002**), *"New Powder Diffraction File (PDF-4) in Relational Database Format: Advantages and Data-mining Capabilities,"* Acta Crystallographica Section B: Structural Science 58 (3), 333-337.

Kaduk, J. A. (**2009**), *"A Rietveld Tutorial – Mullite,"* Powder Diffraction, 24(4), 351-361.

Kaduk, J. A. and Reid, J. (**2011**), *"Typical Values of Rietveld Instrument Profile Coefficients,"* Powder Diffraction, 26(1), 88-93.

Le Meins, J. M., Cranswick, L. M. D. and Le Bail, A. (**2003**) *"Results and Conclusions of the Internet Based Search/Match Round Robin 2002,"* Powder Diffraction, 18(2).

Nusinovici, J., and Beard D. W. (**1988**), *"Specific Data Handling Techniques and New Enhancements in a Search/Match Program,"* Adv. in X-ray Anal., 32, 531-538.

Nusinovici, J. and Bertelmann, D. (**1992**), *"Practical Determination of the Acceptable 2 Theta Error for Non-Ambiguous Identification of Pure Phases with Diffrac-AT,"* Adv. in X-ray Anal., 36, 327-332.

Nusinovici, J. and Winter, M.J. (**1993**) *"DIFFRAC-AT Search: Search/Match Using Full Traces as Input,"* Adv. in X-ray Anal., 37, 59-66.

Reid, J., Crane, D., Blanton, J., Crowder, C., Kabekkodu, S., and Fawcett, T. (**2011**), *"Tools for Electron Diffraction Pattern Simulation for the Powder Diffraction File,"* Microscopy Today, January, 38-42.

Scardi, P., Leoni, M., and Faber, J. (**2006**), *"Diffraction Line Profile from a Disperse System: A Simple Alternative to Voigtian Profiles,"* 21 (4), 270-277.

CASE STUDY: Intelligent RIR – Better Quantitative Analyses

The Reference Intensity Ratio method (RIR) has been used for quantitative phase identification for decades. The intensity of a diffraction peak profile is a convolution of many factors, one of which is the concentration of the analyte (species being measured). By using the RIR method, ratios scaled to a common reference are used in the experiment. The assumption is that all the factors except concentration of the analyte are ratioed and reduced to a constant. By using ratios and measuring peak areas, the RIR method can be used to determine concentrations.

The "Intelligent" RIR method takes advantage of the editorial work that has been performed during quality reviews of reference data published in the Powder Diffraction File®. For example, all entries in the Powder Diffraction File® are given a Quality Mark, Status and Ambient/Non-Ambient designation based upon editorial review. The quality mark is a measure of the consistency between the crystallography, diffraction pattern, and the chemistry of the reference material. The quality mark review contains hundreds of checks, and if all the data are self-consistent, a quality mark of star or "S" is given to the entry. Approximately 20% of published inorganic data have the highest star quality mark. Status marks are Primary, Alternate and Deleted. Primary references are selected by ICDD editors as being the best and most representative data for a material. A low quality reference can still be a primary reference if it is the only reference for a specific material in the database. The ambient/non-ambient designation denotes if a material was measured at non-ambient temperature, pressure or both. At non-ambient conditions, molecules and unit cells can expand or contract resulting in completely different RIR values relative to ambient conditions. Fundamentally, the "Intelligent" RIR option uses the Quality Mark, Status and Ambient/Non-Ambient designation to provide the user with the best available I/I_c value to use in their analysis.

The following analysis uses **Sleve+** in demonstrating "Intelligent" RIR. We deliberately used a complex multiphase specimen to show the advantages, strengths, and weaknesses of the "Intelligent" RIR method.

GOM	PDF #	QM	Status	Coords	I/Ic	Compound Name	Mineral Name	Chemical Formula
6272	00-022-1560	C	P			Ascorbic acid		C6 H8 O6
6221	04-007-8659	I	P	✓	3.2	Calcium Carbonate	Calcite, syn	Ca (C O3)
6207	00-009-0080	I	P	✓		Calcium Hydroxide Phosphate	Monetite, syn	Ca P O3 (O H)
6105	00-022-1536	I	P		1.1	Ascorbic acid		C6 H8 O6
6018	04-012-0489	I	P	✓	3.45	Calcium Carbonate	Calcite	Ca (C O3)
5901	00-047-1743	C	P	✓		Calcium Carbonate	Calcite	Ca C O3
5705	00-023-1730	B	P			Iron fumarate		C4 H2 Fe O4
5665	04-002-3657	I	P	✓	6.07	Potassium Chloride	Sylvite, syn	K Cl
5429	00-005-0586	S	P	✓	2.0	Calcium Carbonate	Calcite, syn	Ca C O3

Figure 1. Sleve+ display panel showing candidate phases with the addition of I/I_c column to use in phase selection for the RIR method.

This is the analysis of a vitamin pill. Several steps were taken to improve identification. In the first step, a subfile of primary, pharmaceuticals and excipients was chosen. In the second step, I/I_c was moved up in the analysis panel of **Sleve+** using the Preferences option in the software. The

first nine candidate phases are displayed in Figure 1 with five different compounds. All of the compounds are in the specimen, but as shown in the display not all compounds have I/I$_c$ values. For example, the reference PDF 00-022-1536 would be selected over reference 00-022-1560 because of the availability of an I/I$_c$ even though the other phase has a higher GOM (better fit to the experimental data).

	#		PDF #	Compound Name	I Ratio	I %	I/Ic	Est Wt %
	1	✓	04-007-8659	Calcium Carbonate	0.276	10	3.2	8
	2	✓	00-022-1536	Ascorbic acid	0.306	11	1.1	25
	3	✓	04-002-3657	Potassium Chloride	0.747	27	6.07	11
	4	✓	04-009-3755	Hydrogen Calcium Phosphate	0.291	10	0.71	36
	5	✓	00-045-0946	Magnesium Oxide	0.192	7	1.0	18
	6	✓	04-004-2931	Zinc Oxide	0.087	3	4.12	2

Phases (7)

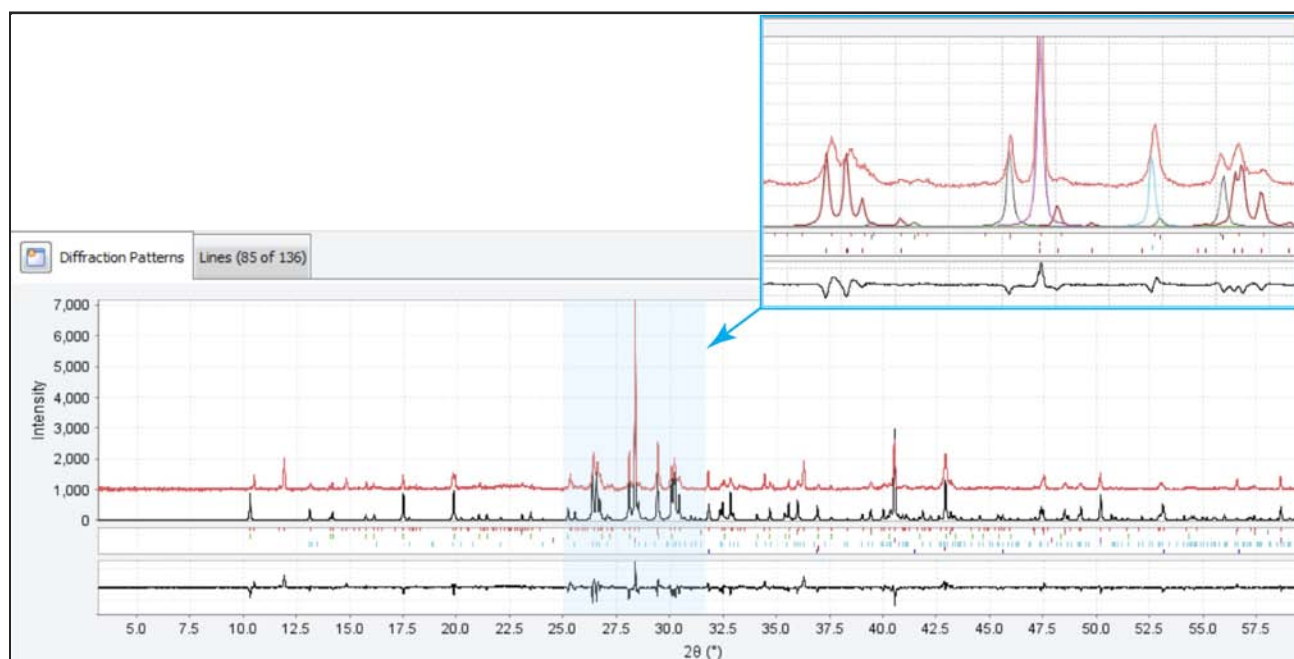*Figure 2. Identification and estimated concentration determined by the RIR method.*

Figure 3. Plot showing the experimental data in red. A scaled combined 6-phase analysis in black and the residual pattern at the bottom. The expanded view on top shows the contibutions of different phases to clusters between 25 and 32 degrees 2θ.

In Figure 2 there are six identified phases and their respective intensity contributions, I/I$_c$ values, and an estimated weight percent calculated by the RIR method using these parameters. For the exact calculation technique, please see the tutorial, "Quantitative Analysis: The Reference Intensity Ratio Method". **http://www.icdd.com/resources/tutorials/pdf/Quantitative%20Analysis%20RIR.pdf**

Figure 3 shows the graphics diagnostic panel of **Sleve+** where these six phases are combined, scaled, and compared to the experimental data. Figures 2 and 3 are part of the diagnostic display panels in **Sleve+** and are usually located on the bottom left and bottom right panels of the diagnostic screen, respectively. Figure 4 shows the RIR analysis if the "Intelligent" RIR option is applied by selecting this option in the "Preferences" panel.

Phases (7)

	#		PDF #	Compound Name	I Ratio	I %	I/Ic	Est Wt %
	1	✓	00-022-1536	Ascorbic acid	0.306	11	1.1	28
	2	✓	00-045-0946	Magnesium Oxide	0.192	7	3.28*	6
	3	✓	04-002-3657	Potassium Chloride	0.747	26	6.07	12
	4	✓	04-007-8659	Calcium Carbonate	0.276	10	3.23*	9
	5	✓	04-009-3755	Hydrogen Calcium Phosphate	0.291	10	0.68*	42
	6	✓	01-070-8070	Zinc Oxide	0.184	6	5.41*	3

Figure 4. "Intelligent" RIR analysis of the same data shown in Figures 1 through 3. The starred I/I$_c$ values denote those that have been replaced by quality analysis criteria.

		PDF-4+	
	RIR	**Intelligent RIR**	**Conc Normalized**
Calcium Hydrogen Phosphate	36	42	42
Potassium Chloride	11	13	18
Ascorbic Acid	25	28	15
Calcium Carbonate	8	9	15
Magnesium Oxide	18	6	8
Zinc Oxide	2	3	2

Table 1. Concentration calculations from RIR as compared to listed ingredients.

In the "Intelligent" RIR example, four I/I$_c$ values were replaced for the phases ZnO, MgO, CaCO$_3$ and Calcium Hydrogen Phosphate. In the case of ZnO, the original I/I$_c$ in PDF 04-004-2931 was calculated from a prototype (QM = P) phase extracted from a publication. This value was replaced by a value from a star quality reference. The median I/I$_c$ value from 16 star quality references is 5.35 (0.14). The Calcium Hydrogen Phosphate reference, 04-009-3755, is a single crystal reference with an "I" quality mark denoting errors in the crystallographic determination, so the I/I$_c$ value was replaced by a star quality reference. The MgO reference from an experimental determination was replaced with an I/I$_c$ from a star quality single crystal determination.

The net result of the standard RIR and "Intelligent" RIR can be compared to a normalized concentration from the list of ingredients, shown in Table 1. The concentrations of three replaced I/I$_c$ values moved closer to the listed concentrations. The calculation for Magnesium Oxide is dramatically influenced. An examination of 68 determinations of MgO I/I$_c$ from high quality powder and single crystal experiments would show the original value in PDF 00-045-0946 to be an extreme outlier, as shown in Figure 5. In both MgO and Calcium Hydrogen Phosphate, the replacement I/I$_c$ values were from highly reliable star quality primary references.
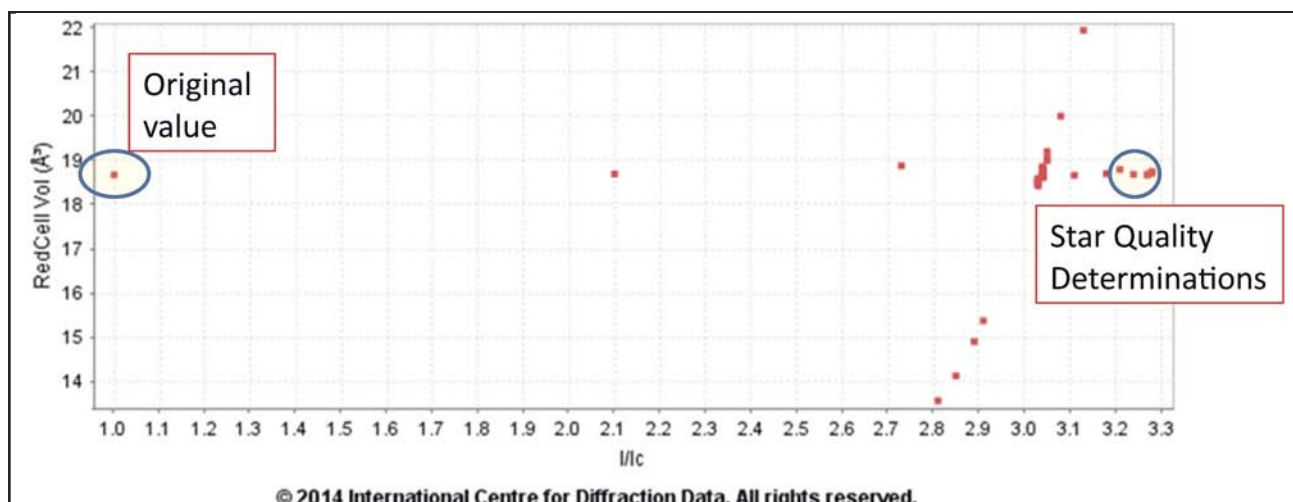
Figure 5. Sixty-eight (68) determinations for I/I$_c$ in Magnesium Oxide. The vertical line data contain many points from a pressure series showing compression and expansion in the unit cell. Star quality ambient data are shown in the tight cluster on the right. The "Intelligent" RIR system evaluates the data and intelligently uses the best quality data when appropriate.

This case study also points out some of the advantages and disadvantages of the RIR method. Three additional phases can be identified from the residual peaks in the pattern shown in Figure 3. These phases are Riboflavin, a B vitamin; Fe Fumarate, a soluble form of iron; and the excipient Cellulose Iβ. I/I$_c$ values are not available for these phases so they could not be included in the analysis. There are atomic coordinates in PDF-4+ for Cellulose Iβ and Fe Fumarate so these two phases could be used in a Rietveld refinement. We plan to include calculated I/I$_c$ values in Release 2014. The crowded peak area between 25 and 35 degrees presents scaling problems due to the lack of peak deconvolution used in this analysis. Total pattern methods that use peak deconvolution such as FullProf, GSAS, Rietveld, and FullPat would be expected to give better quantitative results. Alternatively, we have tested this method on simple mixtures, where peak overlap is not an issue and found it to be very accurate.

While the RIR method for complex multiphase samples is semi-quantitative, the "Intelligent" RIR method does improve performance with fast automated analyses.

It should also be noted that the concentration of Ascorbic Acid as determined by this experiment is noticeably higher than what exists in the specimen. This is typical of a phenomenon known as microabsorption. When the crystallite size of the specimen is large enough to approach or exceed the typical penetration depth of X-rays, then the lower-absorbing materials receive and diffract a larger-than-proportional fraction of the incident X-rays. This causes the diffraction pattern from that lower-absorbing material to be artificially more intense. This cannot be corrected via RIR selection — only by reducing the crystallite size of the specimen and re-collecting the data can this be corrected.

Recommended Settings When Using SIeve and SIeve+

The following recommendations are based on years of alpha and beta testing using 3-10 phase experimental mixtures of a wide variety of materials and analyzing the results. As explained in the technical bulletin, each algorithm has strengths and weakness. These recommendations are based on algorithms that work with most data sets most of the time.

KEYS TO SUCCESS

Subfiles: Always Select the Appropriate Subfile

Most analysts know if they are working on a criminal investigation (forensics); cement (cement and hydration products); pharmaceutical drug (pharmaceuticals); or soil sample (minerals). The selection of one of more than 30 classified subfiles and subclasses in the Powder Diffraction File™ (PDF) targets the chemistry of the search, reduces the number of false positives, and dramatically increases the chances of a correct identification.

Any search in PDF-2 or PDF-4 can be used to make a user defined subfile. You can create subfiles based on content type (i.e., minerals), elemental composition, physical properties, etc.).

Determining Match Quality: GOM – Goodness of Merit

The GOM is a very general scoring algorithm based on comparing multiple d-spacings in the experiment and reference materials. It is a general, all purpose scoring method that primarily depends on d-spacing accuracy and precision in both the experiment and reference. Good quality input data will produce good results. *This method works best in conjunction with a graphical interface so that the user can see the GOM score (d-spacing match) and visually check the intensity match through the graphics display.* The GOM scoring method is based on the analysis of crystalline materials. Alternative methods should be deployed with non-crystalline materials.

2nd Pass Filter (ON)

This should be switched "ON". The filter looks for major (most intense) peaks in the reference that are not found in the experimental data. It reduces the number of false positives. By toggling this filter "ON", you should see that the number of candidate phases drops significantly. The total number also depends on your GOM threshold.

If you work with pharmaceutical materials, you may want to toggle this filter (ON/OFF) and compare the results. Orientation is a very common problem with pharmaceuticals. The 2nd pass filter may eliminate an oriented material because of the intensity variations caused by the orientation.

RECOMMENDATIONS

Raw Data Treatments: Background Subtraction (yes), Data Smoothing (no), 2nd Derivative Peak Identification (yes).

Background subtraction can often be an art, but it is essential for measurements of trace phases, all quantitative analyses, and crystallinity measurements. Data smoothing is a means

of correcting poor quality data, but it can also produce artifacts in the data (use only when needed). The second derivative peak finding methods is very successful in the majority of cases.

Primary Patterns, Ambient Patterns: Select

Patterns are chosen by experts; if possible, select both primary patterns and ambient patterns. Primary patterns remove redundant alternates and poor quality patterns. Ambient pattern selection removes variable temperature/pressure patterns, which should not be present in experiments run at ambient conditions. Removal of inappropriate data reduces the chances of having a false result.

Search/Match Algorithm: Hanawalt

The Hanawalt method is a proven method in use for ~75 years on the identification of crystalline materials. It works on a very wide range of specimens and is the best place to start with an unknown analysis.

GOM Threshold: High at the Start (4000) of an Analysis and Low at the End (0-1000)

This threshold setting determines the number of candidate phases in your candidate list and is based on d-spacing matches of 8 lines, for a maximum score of 8000. If you have high quality data you should get many candidates with scores >4000. This usually represents high concentration phases where there are many peaks available for matching. However, once the major intensity phases are identified you may only have a couple of unidentified peaks. To determine these minor or trace concentration phases, the GOM has to be lowered. If you have a pattern with many peaks, but all candidate GOM's are low (2000 and below), you may have a specimen transparency or displacement problem. You can check this, and then make adjustments within your graphic display. Choosing a low GOM for you initial analyses will result in a large number of candidates and slower response times.

Contact Information

phone	610.325.9814
toll free	866.378.9331
	(U.S. & Canada)
fax	610.325.9823
email	info@icdd.com
web	www.icdd.com
registration	registration@icdd.com
sales	sales@icdd.com
technical support	support@icdd.com

www.icdd.com

4 Easy Ways to Order

online	www.icdd.com/products
email	sales@icdd.com
fax	610.325.9823
mail:	International Centre for Diffraction Data
	12 Campus Boulevard
	Newtown Square, PA
	19073-3273 U.S.A.

Please send us a sales order via fax, email or mail. ICDD requires the name of the end user on all orders. Use product codes in your order, if possible. A convenient order form can be found on our website.

E-commerce

From the Powder Diffraction File to ICDD Clinic registration, ordering ICDD products and services has never been easier!

Pricing

The prices listed herein supersede all others previously published and are subject to change without notice. ICDD reserves the right to discontinue any product at its discretion. Visit our website at www.icdd.com or contact our Customer Service Department for the most recent product information and pricing.

Return Policy

Products may be returned within 30 days from the invoice date for a full refund or credit (excluding shipping/handling). Products must be returned in original condition to qualify for a refund or credit. Registered products do not qualify. Please complete a Return Merchandise Authorization (RMA) form to receive approval prior to returning product. Visit www.icdd.com/products/RMA.doc to download the RMA form. Defective products will be replaced promptly, but must be returned for testing.

Warranty Policy

Customers' sole remedy for any defective Licensed Product shall be replacement of the Licensed Product, which right must be exercised by written notice given within 30 days of purchase. Defective products will be replaced promptly, but must be returned for testing.

Payment Terms

Wire transfers, credit cards (Mastercard, Visa, American Express) and checks in USD$ and drawn on U.S. banks are acceptable forms of payment. Letters of credit will not be accepted. Payment is due within 30 days from invoice date. Finance charges of 1.5% per month (annual percentage rate of 18%) will be applied to overdue balances.

The ICDD Online Store Offers:

- Secure transmission using SSL encryption
- Customer on-line order history
- Online order tracking
- Email order confirmation
- Email shipping notification
- Customer satisfaction

Licensing

ICDD licenses the use of the Powder Diffraction File and software. Licensing for the PDF-2, PDF-4 product line, and associated software (SIeve and SIeve+) are packaged with the CD, DVD and/or dongle. ICDD's product licenses are non-transferable and intended for use on a single computer. It is expressly prohibited to access the database over a network or access the database using desktop sharing or screen sharing software.

Please refer to the sales descriptions and product policies for further information on licensing. Licensing for other products requires users to contact ICDD's Customer Service Department.

Shipment

All orders shipped by ICDD are F.O.B. Newtown Square, PA, U.S.A. **Import duties and all taxes are not included in pricing and are the sole responsibility of the customer.** Shipment will be made by a commercial carrier chosen by ICDD. For your convenience, shipping charges will be added to all invoices. Shipments can be insured upon request; additional charges will apply.

Orders & Delivery

All orders are handled on a First In, First Out (FIFO) priority queue. Delivery will occur approximately 30 days on items in stock after receipt of completed order.

Note: Customers in Japan, Korea and Russia must order through an ICDD licensed vendor.

ICDD®
INTERNATIONAL CENTRE FOR DIFFRACTION DATA

ISO 9001:2008
QUALITY ASSURED COMPANY
CERT.NR. 11409.1

International Centre for Diffraction Data
12 Campus Boulevard
Newtown Square, PA 19073-3273 U.S.A.
www.icdd.com
t: 610.325.9814
f: 610.325.9825
Toll-free 866.378.9331