ICDD Headquarters, Newtown Square, PA, USA

POWDER DIFFRACTION FILE: RECENT DEVELOPMENTS IN QUALITY CONTROL

Soorya N Kabekkodu (suri), Timothy Fawcett

International Centre for Diffraction Data

Newtown Square PA, USA

DATABASE OVERVIEW

- Quality Marks
- Database Status of a given entry
- Structural Classification
- Subfile Designation

QUALITY MARKS

- Quality marks are essential while working with larger database with similar diffraction patterns
- Data validation and the quality mark assignments is the most important step in the editorial process

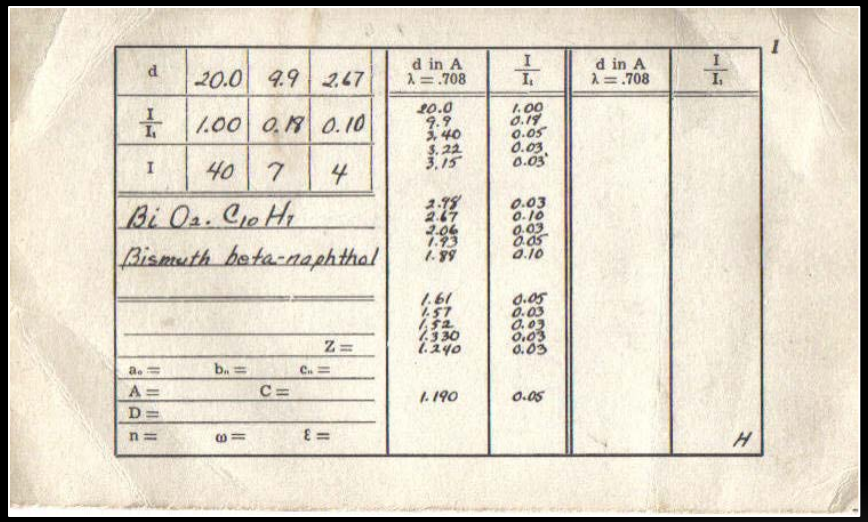
DATA QUALITY

1941 Data Entry by Don Hanawalt

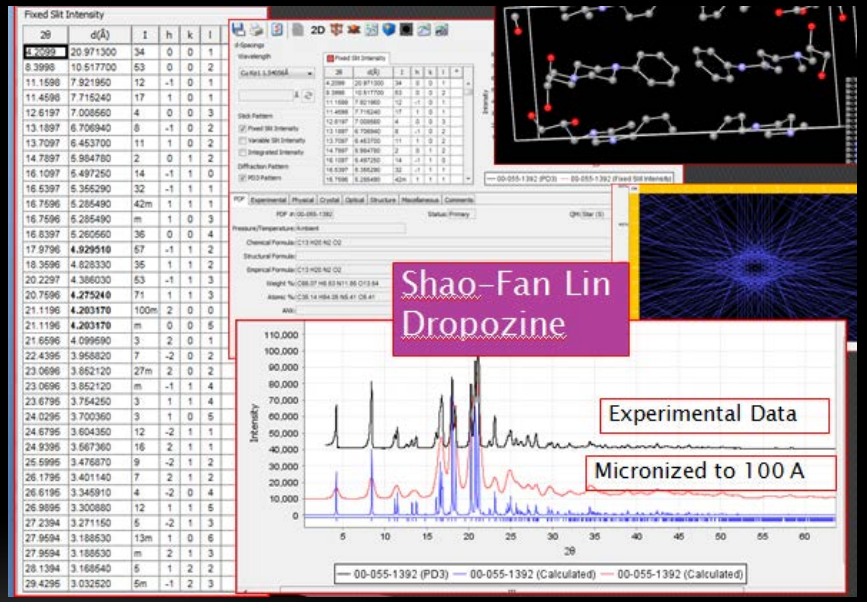
Set 1, card 1

2005 Data Entry by Shao-Fan Lin

Set 55, entry 00-055-1392



d	20.0	9.9	3.47	d in A λ = .708	I/I1	d in A λ = .708	I/I1
I/I1	1.00	0.18	0.10	20.0	1.00		
I	40	7	4	9.9	0.19		
				3.40	0.05		
				3.22	0.03		
				3.15	0.03		
Bi O2 . C10 H7				2.78	0.03		
Bismuth beta-naphthol				2.67	0.10		
				2.06	0.03		
				1.93	0.05		
				1.89	0.10		
				1.61	0.05		
				1.57	0.03		
				1.52	0.03		
				1.330	0.03		
				1.240	0.03		
Z =							
a0 =	b0 =	c0 =					
A =	C =			1.190	0.05		
D =							
n =	ω =	ε =					



Data quality improves with time, historical data were typically recorded on photographic film with visual estimates for peak positions and intensities. Modern data are typically recorded with an automated diffractometer where digital patterns are recorded using position sensitive detectors (both 1D and 2D) and processed by automated programs with embedded data corrections for instrumental and specimen errors.

Powder Diffraction File

- Contains (Inorganic+Organic) more than 760,000 entries
- 214,882 with atomic coordinates
- All the entries have quality mark indicating the reliability
- Powder Diffraction File is the ONLY crystallographic database that classifies entries based on quality mark
- Assignment of structural prototype and their classification
- Entries are classified based on subfiles
- Data validation, quality mark assignments, structure prototyping are the most important step in the editorial process

EDITORIALIZED DATABASE FILTERS

- Database contains multiple experimental determination of a given phase on account of
  - Different experimental condition
  - Publication from different research groups (these materials are obviously of high research/industrial interest)
- Editorially assigned quality marks and database status flags are vital in optimizing queries dealing with phases having multiple determination

QUALITY MARK (QM) TYPES

- Experimental patterns
- Calculated Patterns (based on experimentally determined crystal structures)

QM FOR EXPERIMENTAL PATTERNS

- Star (Well characterized chemically and crystallographyically, No unindexed lines,$\Delta 2\theta \leq 0.03^{o}$)
- R (d values from whole pattern fitting, like Rietveld, Le Bail refinement)
- I (Well characterized chemically No unindexed strong lines,$\Delta 2\theta \leq 0.06^{o}$)
- B (Do not meet the criteria for *, I)
- O (Poorly characterized, with editorial comment explaining the reason)
- C (author calculated d values)
- H (Hypothetical)
- M (Minimally acceptable non crystalline pattern)
- G (Good non crystalline pattern, usually has additional characterization other than XRD)

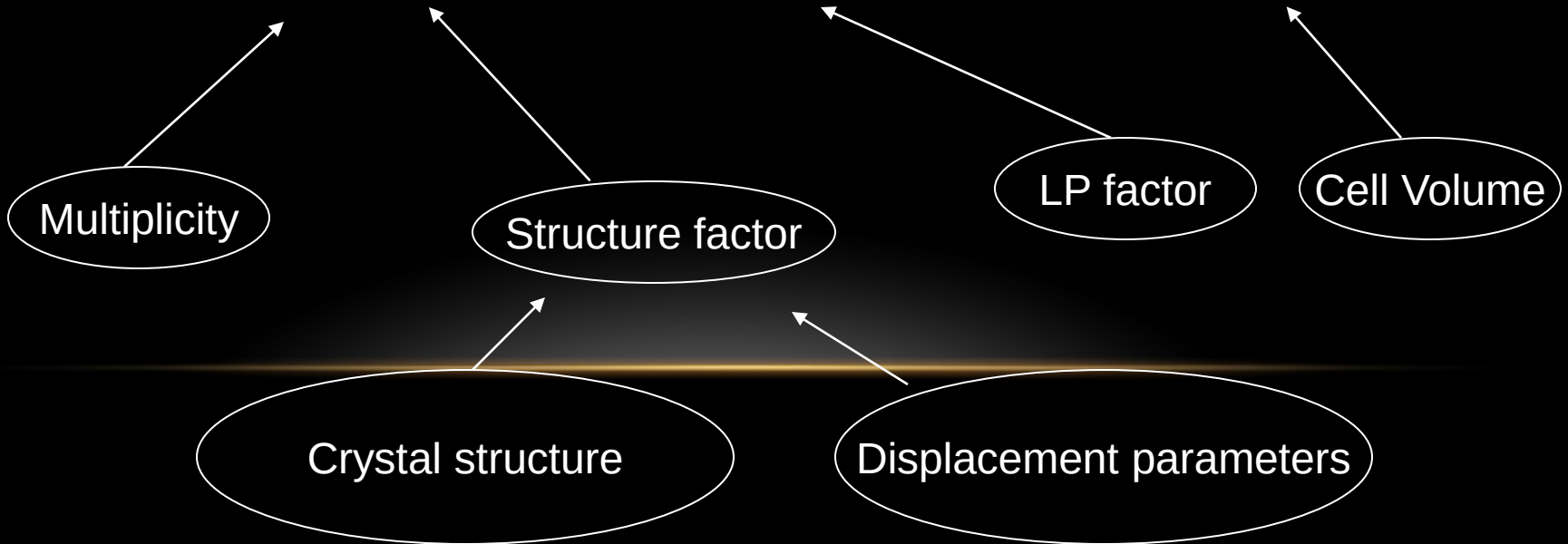
QM FOR EXPERIMENTAL PATTERNS

- Each and every pattern editorially reviewed including visual comparison of raw data vs d-I list
- Undergoes more than 100 data validation checks before gets published

CALCULATED PATTERNS

If we know the crystal structure, we can calculate the diffraction pattern using the equation

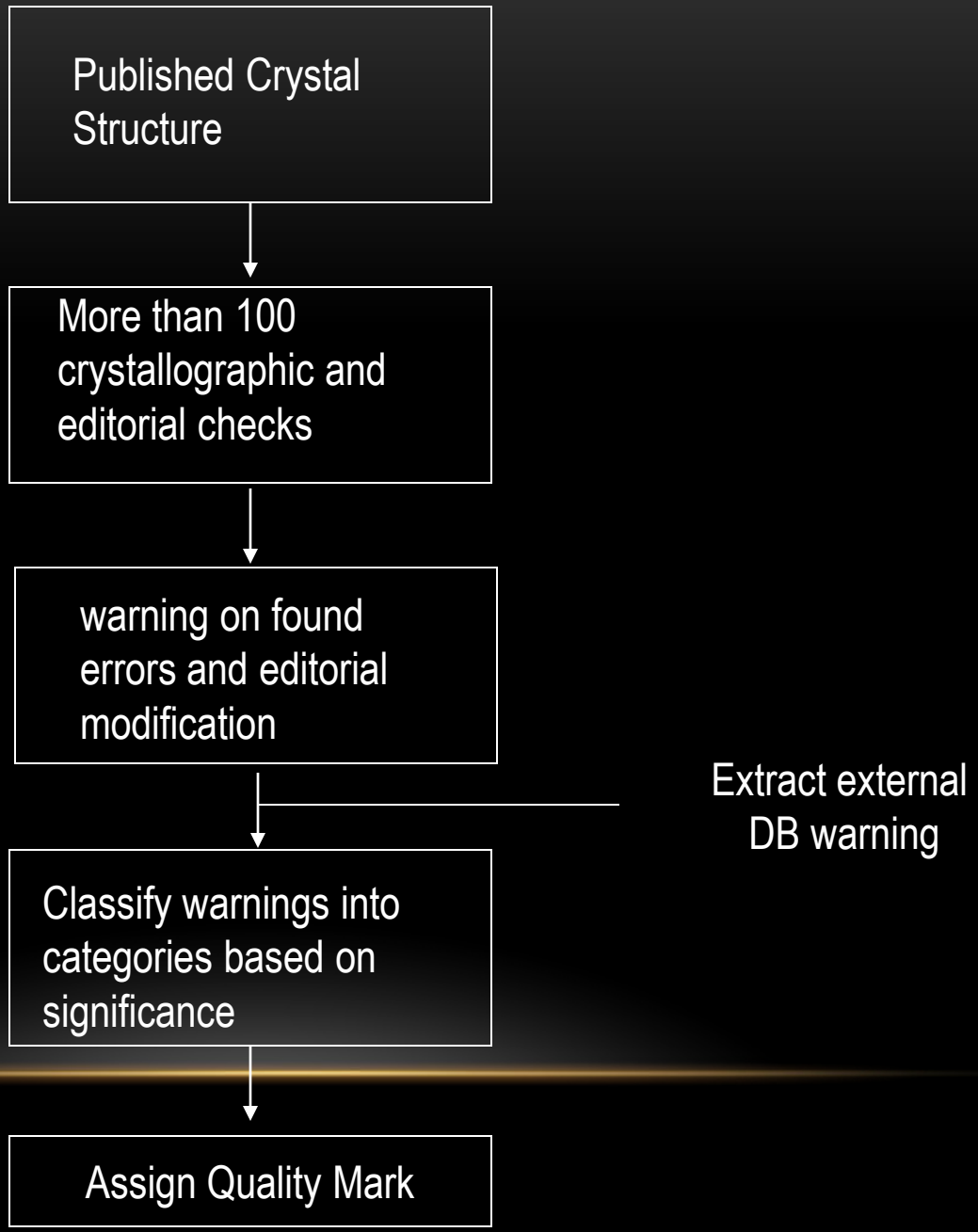
$$I_{hkl} = M_{hkl}\left|F_{hkl}\right|^2\left(1+\cos^2(2\theta)\cos^2(2\theta_m)/\sin^2\theta\cos\theta\right)_{hkl}/V^2$$



It is extremely important to make sure that the crystal structure used for the calculation is correct. In fact it is the rate determining step in the editorial process

THE METHOD

- Editor classifies and flags warnings
  - Minor warning (W1)
  - Significant warning (W2)
  - Major warning (W3)
- QM will be assigned SQL query based on type of warning code(s).


Published Crystal Structure
More than 100 crystallographic and editorial checks
warning on found errors and editorial modification
Extract external DB warning
Classify warnings into categories based on significance
Assign Quality Mark

CALC. QM NOTATIONS

Category	QM
No Warning	*
Minor Warning	I
Significant warning	B
Assigned structure (Prototype)	P
Hypothetical	H
Major warning	O

Overall quality

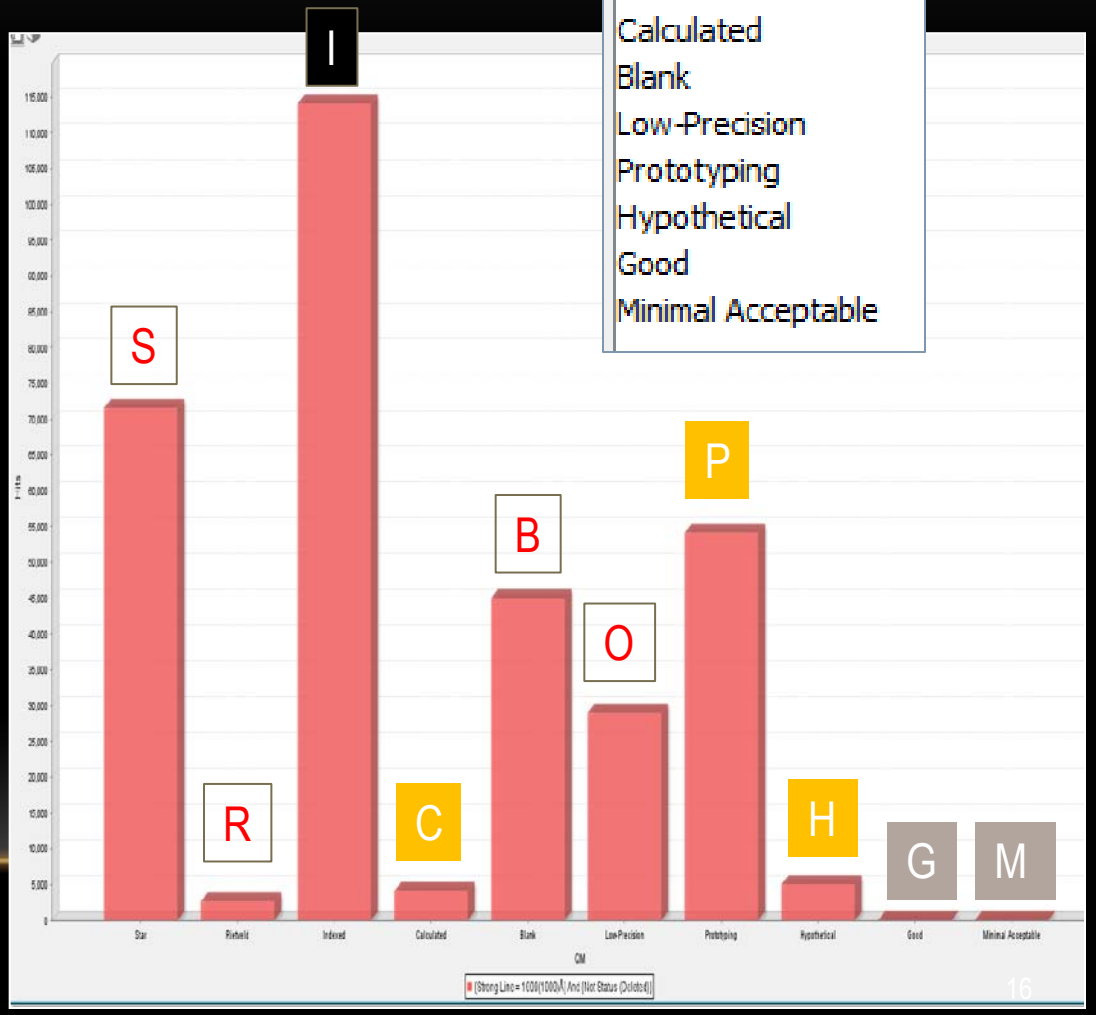
This is the quality mark distribution for

340,653 entries published in PDF-4+ Release 2013. Quality is defined by the ability to produce an accurate powder pattern with known and consistent chemistry, physical properties, and unit cell parameters. The ICDD does over 100 separate checks in its quality review.

Star (S), Indexed (I), and Blank (B) marks are in decreasing order of quality. In these cases there is a unit cell and calculated d-spacings which can be compared to experimental d-spacings. Low precision data (O) tend to be historical in nature where d-spacings and intensities were visually estimated resulting in low precision. Rietveld (R) quality marks result from Rietveld refinements where there is a strong agreement with the structure and powder patterns resulting in high quality.

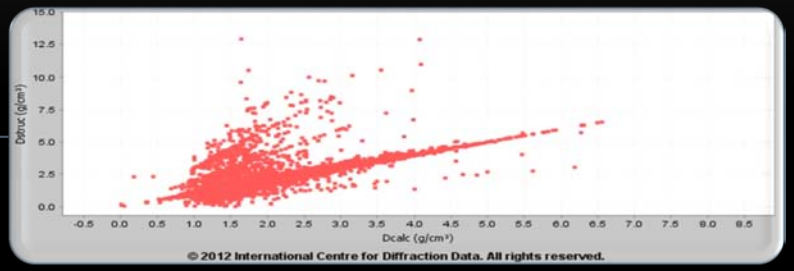
Calculated (C), Prototype (P) and Hypothetical Data (H) all result from calculation *and do not have corresponding experimental data.* The quality of these data are more difficult to ascertain. The user should be cautious using these data.

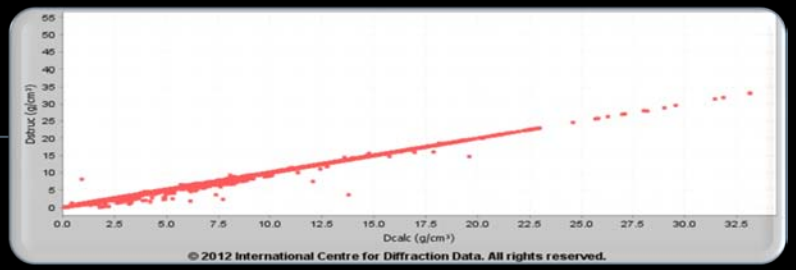
Good (G) and Minimal Acceptable (M) are new quality marks being used for non-crystalline or partially crystalline solids. The descriptors (G or M) describe the quality of supporting analytical data and the reproducibility of the reference powder pattern.

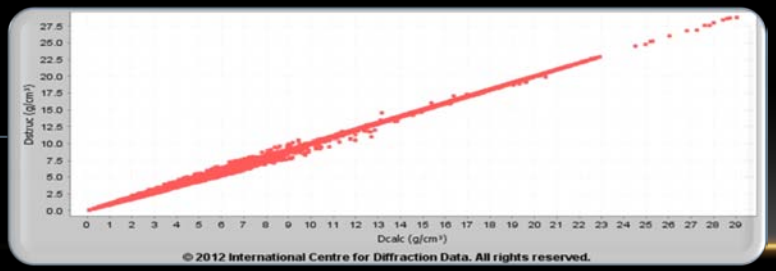


ICDD QUALITY SYSTEMS – STRUCTURAL VS CALC DENSITY

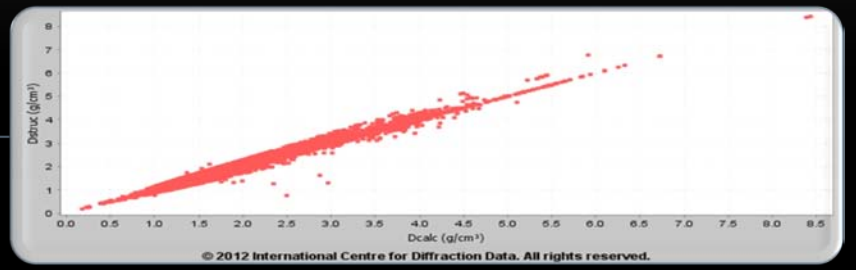
Before Quality Mark

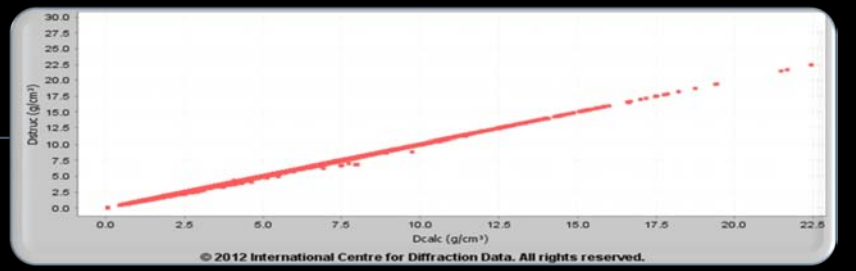


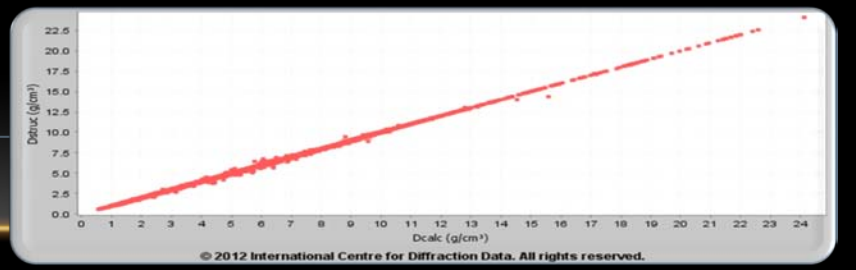




After Quality Mark (2005)



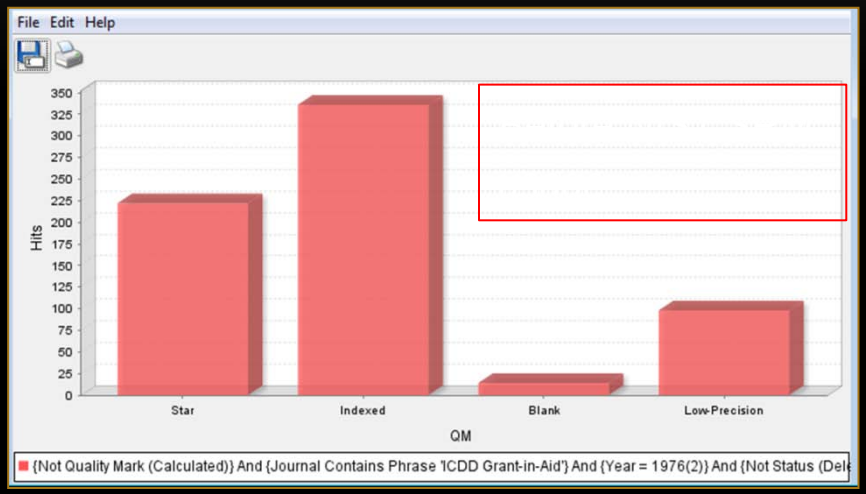




QUALITY IN POWDER DIFFRACTION DATA

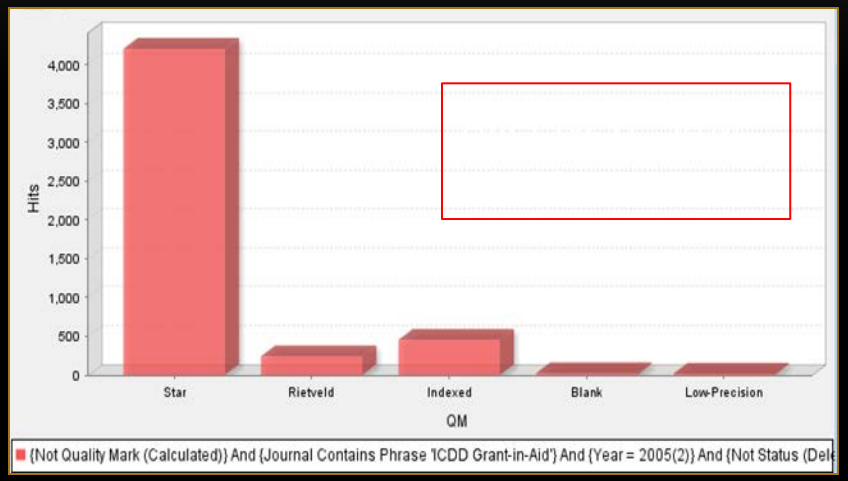
Quality Mark Distribution

ICDD Grantees 1973-1977



Quality Mark Distribution

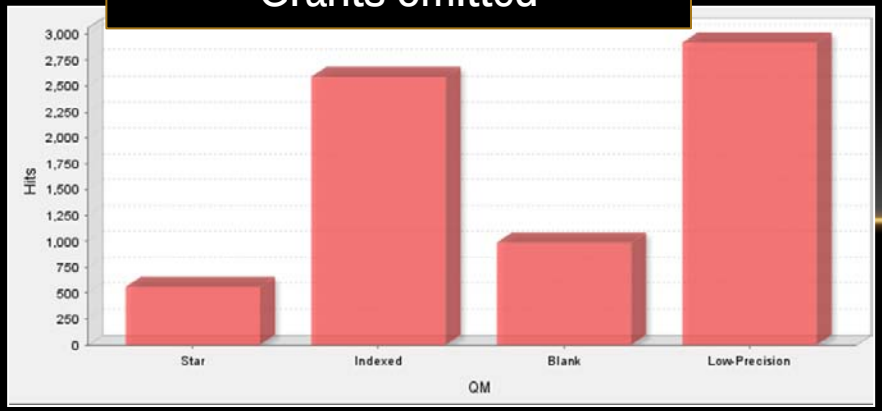
ICDD Grantees 2003-2007



Quality Mark Distribution

All Powder Data 1973-1977

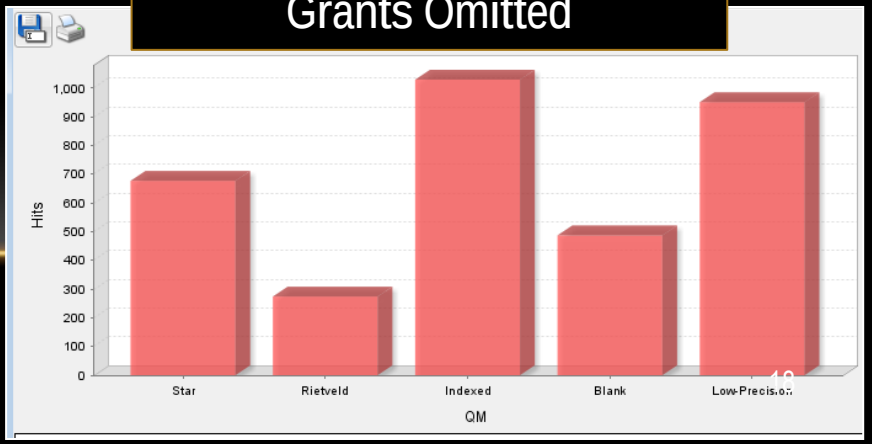
Grants omitted



Quality Mark Distribution

All Powder Data 2003-2007

Grants Omitted

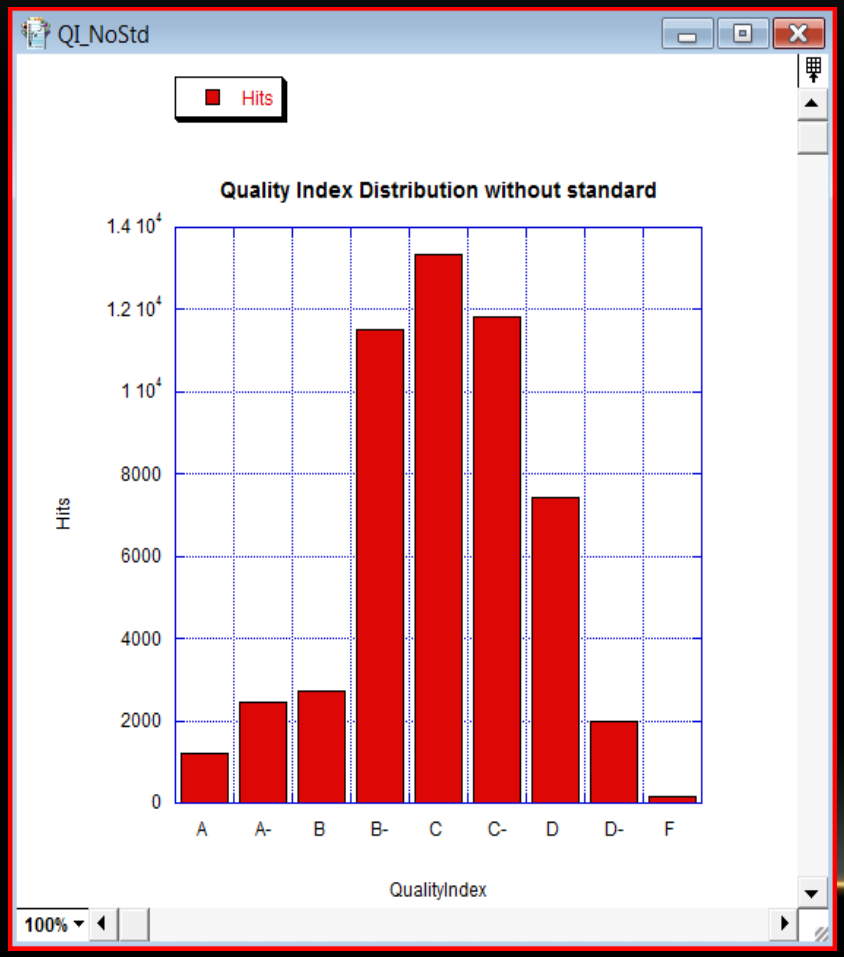


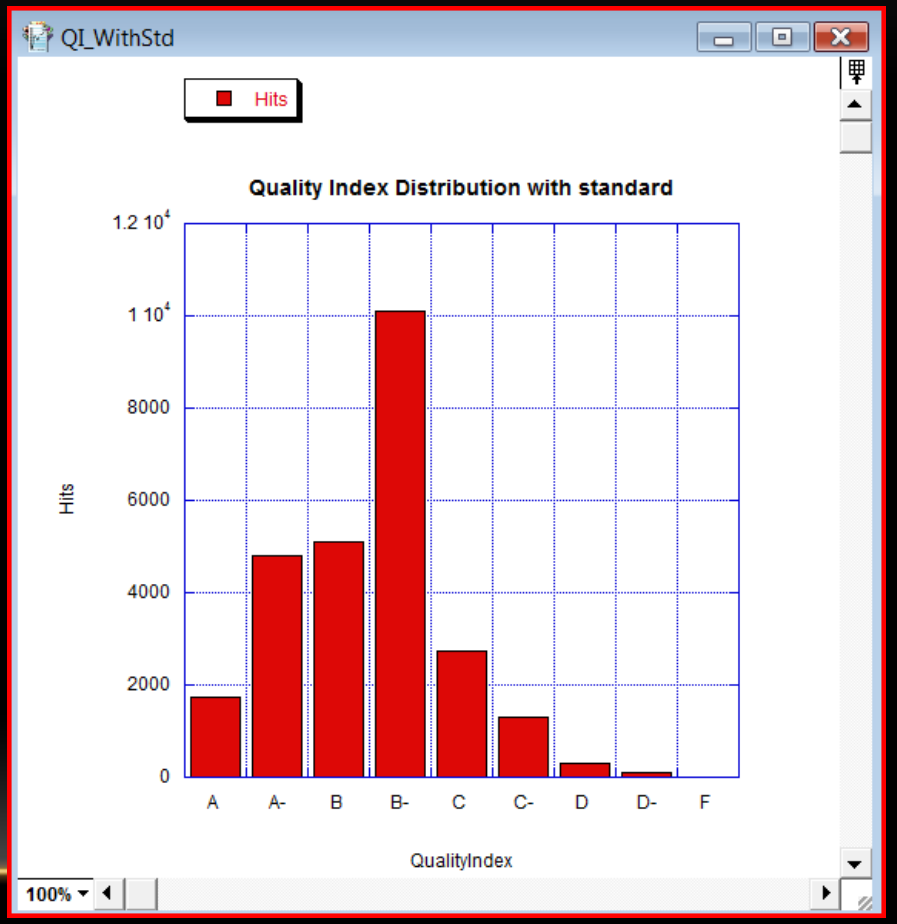
QUALITY INDEX

- Quality Index in Powder diffraction file is a measure of average error (parts per 10^5) associated with unit cell dimensions usually denoted by a letter

0-1	A
2-5	A-
6-10	B
11-50	B-
51-100	C
101-500	C-
501-1000	D
1001-5000	D-
5001-9999	F

QUALITY INDEX DISTRIBUTION COMPARISON





Quality and Phase Identification Results

This is a copy of an ICDD alpha test where we evaluate the ability of the database to identify unknowns using test data sets. In this case, the test data sets were mixtures of vitamin pills, minerals, fertilizers and cements. We record the phase identified, the match score (higher is better) and the reference quality.

The highest quality reference data (S,R) was used in the identification of 44% of the phases even though these data represent 21.7 % of the database. S, R and I quality data were used for 88 % of the identifications.

	Phase	Score	QM		Phase	Score	QM
Test 1	Calcite	6412	S	Test 5	C3S	5393	R
	zincite	5399	S		Anhydrate	4906	I
	Ascorbic Acid	6176	C		CaSO4-0.5 H2O	3354	I
	KCL	5854	I		Brownmillerite	2862	S
	Monetite	5853	I		Dolomite	3243	S
	cellulose	3504	S				
	Brushite	3080	S	Test 6	Muscovite	6977	I
					Chlorite serpentine	5733	I
Test 2	Calcite	6181	I		Quartz	5259	I
	Vit C	5871	I		Rutile	4745	
	KCl	5854	I				
	Monetite	5473	I	Test 7	Urea	5177	S
	zincite	5853	S		Ammonim H Phosphate	4914	P
	cellulose	3708	S		Sulfer	4160	I
					KCl	3898	I
Test 3	C2S	4784	R				
	SiO2	3447	B	Test 8	Calcite	5767	I
	MgO	3128	H		Monetite	5645	I
					Zincite	6110	S
Test 4	quartz	5140	S		Sylvite	5316	I
	Azurite	4577	S		Ascorbic Acid	3963	C
	(Pb) Barite	5708	S		Brushite	3047	S
	Florite	3376	I		Iron Fumarate	1401	R
	Malachite	2268	S		MgO	1400	I
					Cellulose	1466	S

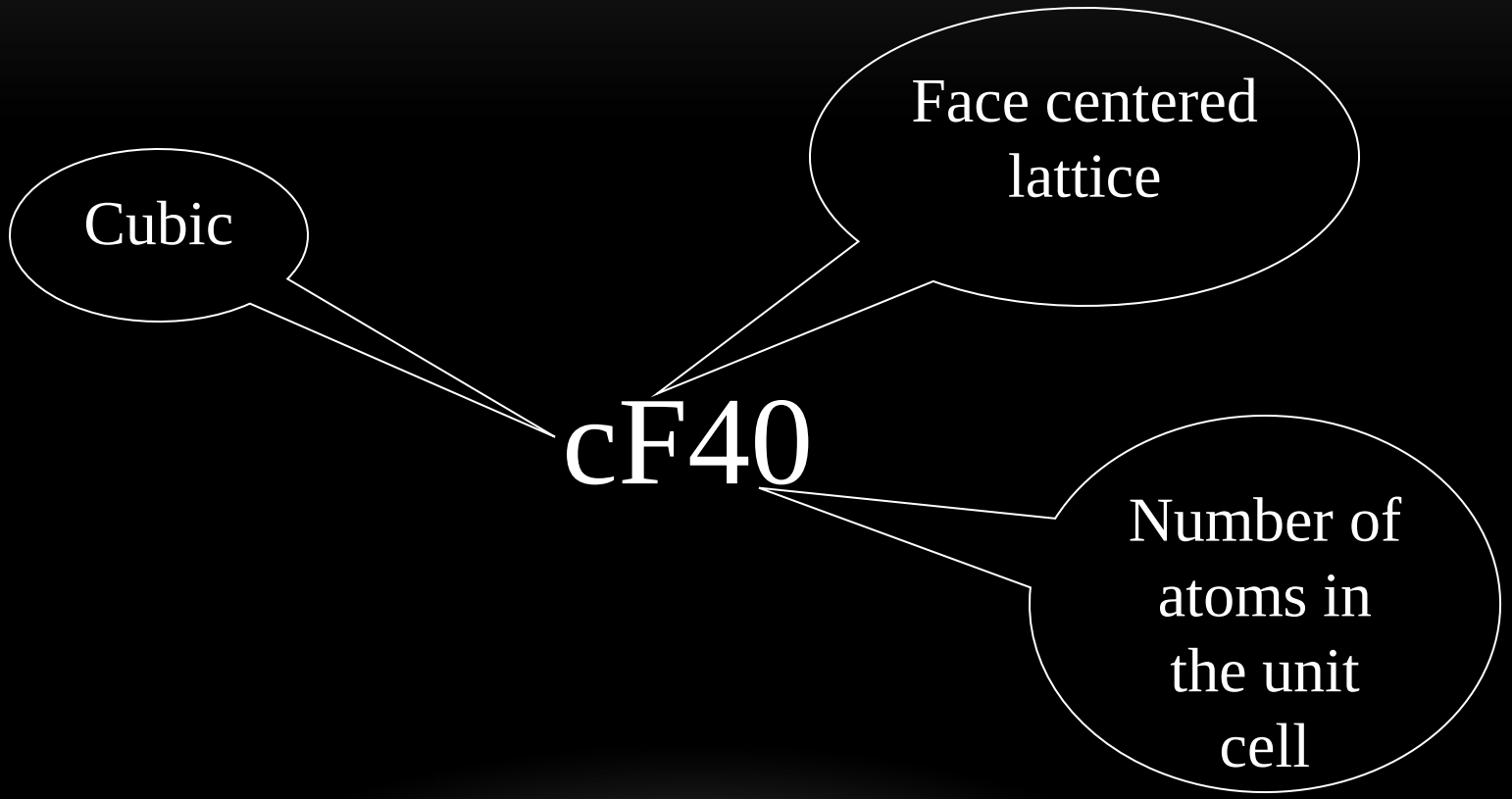
For several consecutive years, similar results have been obtained in alpha tests of ICDD products (PDF-2, PDF-4+, and PDF-4/Organics databases). Higher quality references most frequently match higher quality data produced by modern instrumentation and software.

The averaged test scores have also increased with time as higher quality references replace lower quality references with time, producing higher scores. These tests have also been repeated with major commercial search/match software used by large global manufacturers of XRD equipment.

DATABASE STATUS

- When multiple entries are available, database status is assigned editorially to avoid nearly duplicate result set
- Primary
  - Primary pattern for a given phase usually best quality, room temperature data
- Alternate
  - Alternate to primary. Not necessarily mean of poor quality
- Deleted
  - Unresolved errors
  - Duplicated entry

STRUCTURAL CLASSIFICATION: PEARSON SYMBOL CODE (PSC)



STRUCTURAL CLASSIFICATION: STRUCTURE TYPE NOTATION

- Popular structure type descriptions are
  - Traditional Notation
    - Based on unit cell, Pearson Symbol and chemistry (usually assigned manually by comparing the diffraction patterns)
  - ANX Formula
    - Based on type of ion and their site occupation
    - Example
      - Ca Ti O3 is of ABX3 type
      - Fe3O4 is of AB2X4 type
  - Long descriptive Notation following Parthe's method
    - Based on detailed crystallographic analysis
    - Example Cu3 As,cI64,220 (Structure Type Formula, Pearson Symbol, Space group number)
    - Sensitive to prototype branching based on atomic environment

STANDARDIZATION OF CRYSTAL STRUCTURE

- First proposed by Parthe and Gelato (*Acta Cryst. (1984), A40, 169-183*)
- The method uses standardization parameter

$$\Gamma = \sum_{i=1}^{N} \left(x_i^2 + y_i^2 + z_i^2 \right)^{1/2}$$

N is number of atom sites, x,y,z are fractional coordinates

WHY STANDARDIZED DATA FOR COMPARISON?

CeCu2, Imma, a=4.425,b=7.057, c=7.475

Cu (8h) 0 0.051 0.1648

Ce (4e) 0 0.25 0.5377

Standardization

CeCu2, Imma, a=4.425,b=7.057, c=7.475

Cu (8h) 0 0.051 0.1648

Ce (4e) 0 0.25 0.5377

KHg2, Imma, a=8.10,b=5.16, c=8.77

Hg (8i) 0.190 0.25 0.087

K (4e) 0 0.25 0.703

Standardization

KHg2, Imma, a=5.16,b=8.10, c=8.77

Hg (8h) 0 0.06 0.163

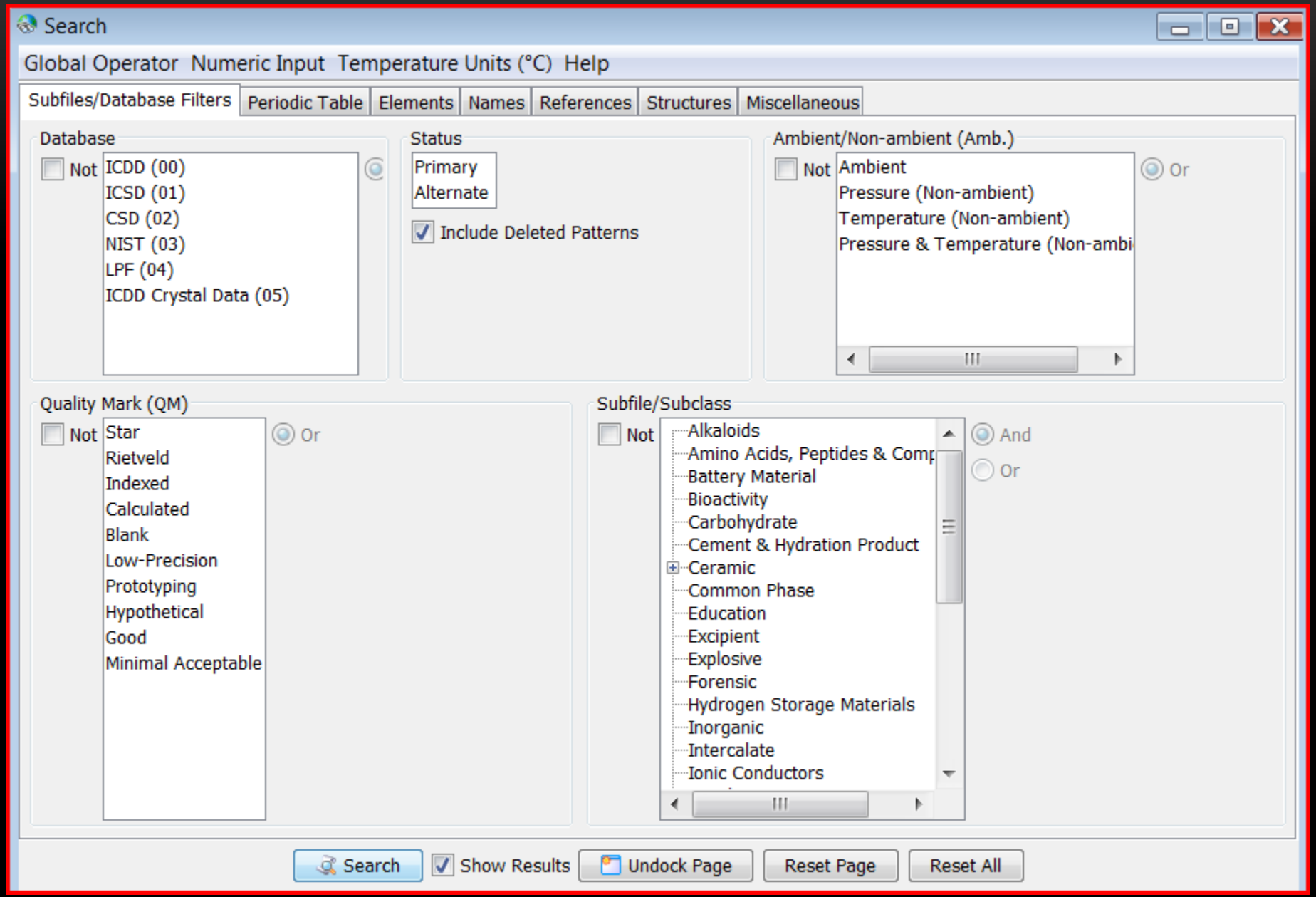
K (4e) 0 0.25 0.547

SUBFILES

- Classified into various categories based on chemistry, properties and application
- Each subfile is defined by set of rules approved by the concerned subcommittee
- Subcommittee members review the file and continuously improvise it
- Examples: Cement, Pharmaceutical, Ion Conductor, Forensic etc

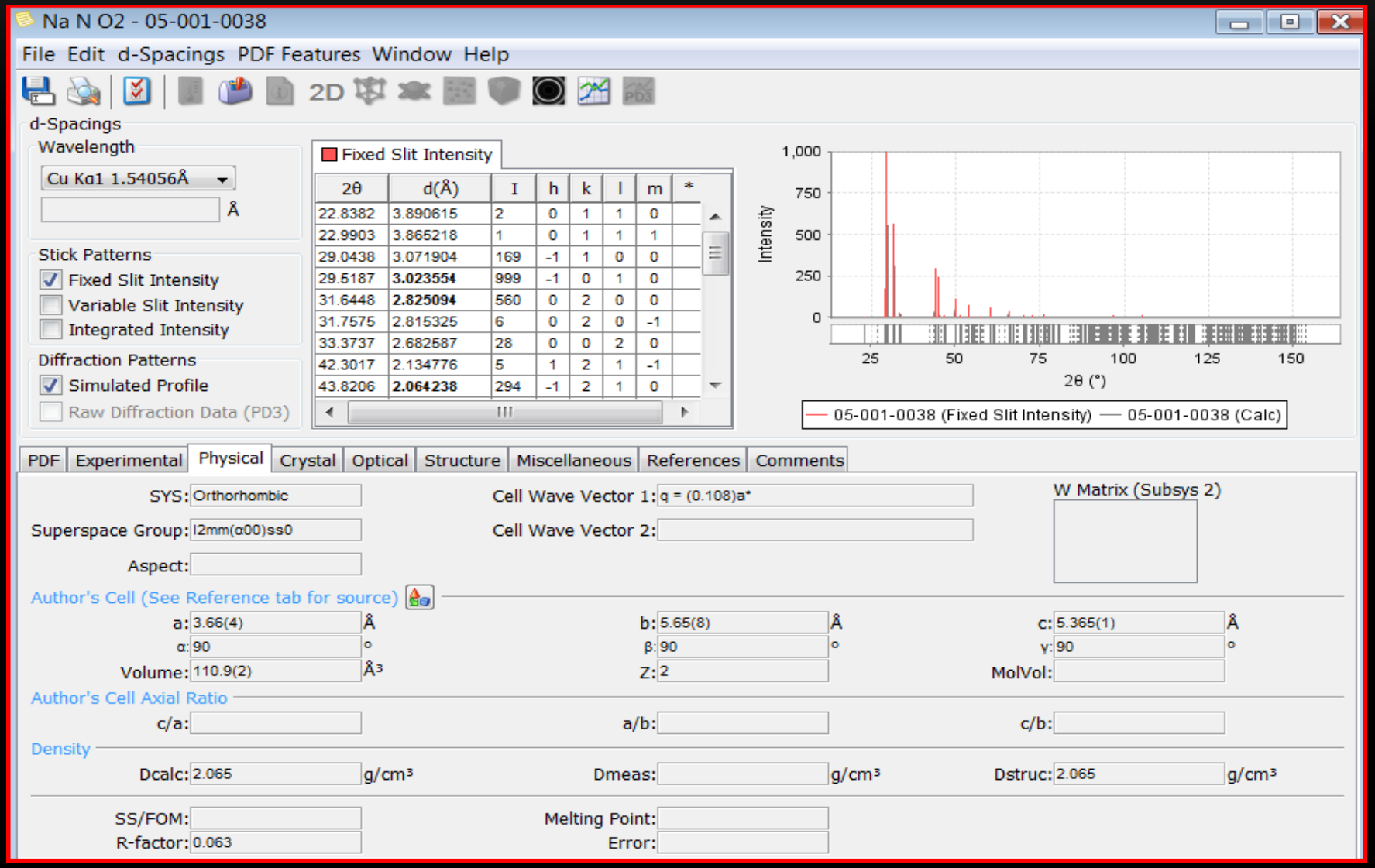
THE POWER OF SUBFILE SEARCHING

- This feature allows you to search a user-specified subset created using the many search criteria of DDView.
- You may ‘limit’ the database to just ceramics, or iron-containing minerals, or cubic perovskites, or pharmaceuticals, or superconductors containing copper, or primary patterns, or . . .
- By focusing the database in such a manner, the number of coincidental ‘near-hits’ is correspondingly reduced and the speed of search/match is correspondingly increased.

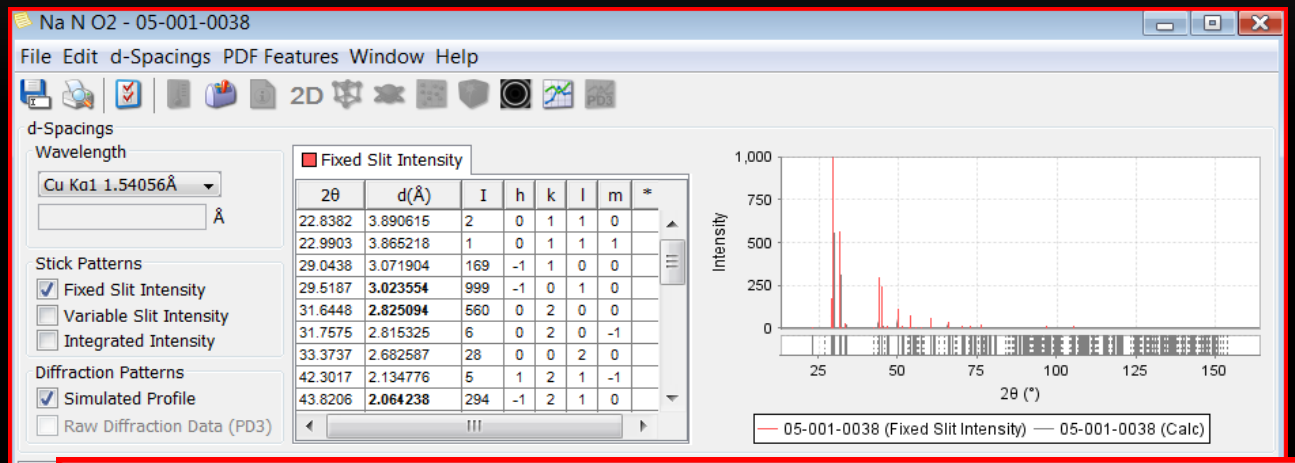

Search
Global Operator Numeric Input Temperature Units (°C) Help
Subfiles/Database Filters
Periodic Table
Elements
Names
References
Structures
Miscellaneous
Database
Not
ICDD (00)
ICSD (01)
CSD (02)
NIST (03)
LPF (04)
ICDD Crystal Data (05)
Status
Primary
Alternate
Include Deleted Patterns
Ambient/Non-ambient (Amb.)
Not
Ambient
Pressure (Non-ambient)
Temperature (Non-ambient)
Pressure & Temperature (Non-ambi
Or
Quality Mark (QM)
Not
Star
Rietveld
Indexed
Calculated
Blank
Low-Precision
Prototyping
Hypothetical
Good
Minimal Acceptable
Or
Subfile/Subclass
Not
Alkaloids
Amino Acids, Peptides & Comp
Battery Material
Bioactivity
Carbohydrate
Cement & Hydration Product
Ceramic
Common Phase
Education
Excipient
Explosive
Forensic
Hydrogen Storage Materials
Inorganic
Intercalate
Ionic Conductors
And
Or
Search
Show Results
Undock Page
Reset Page
Reset All

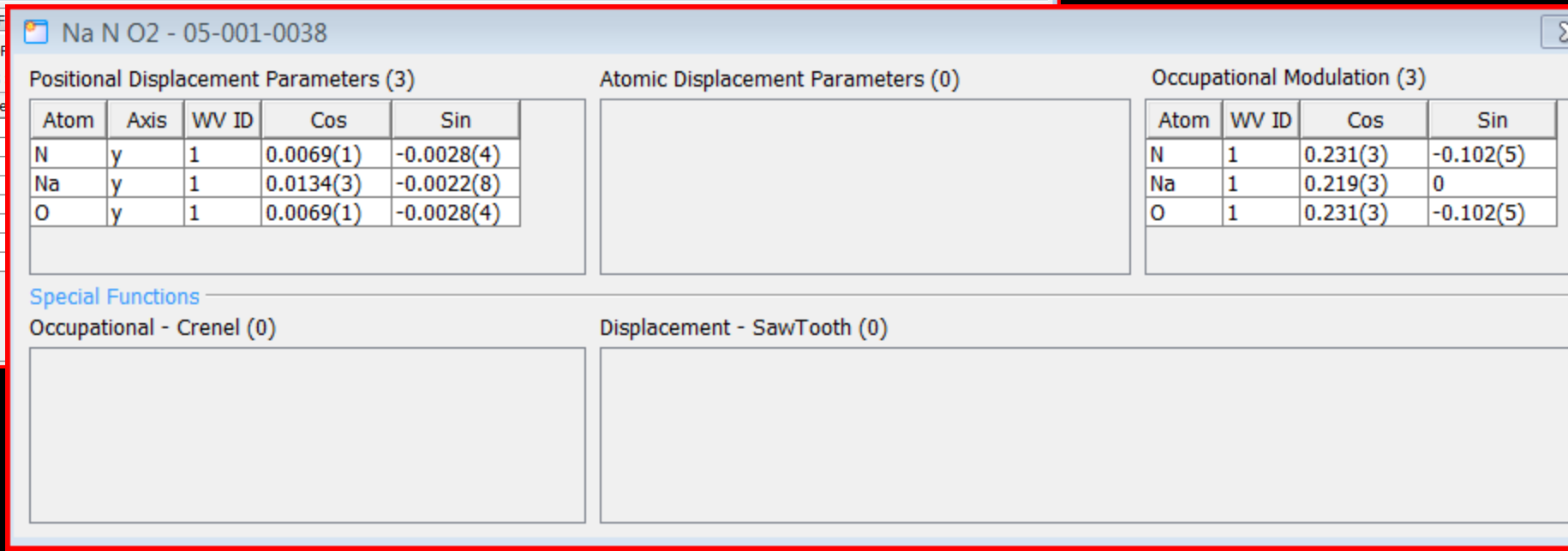
RECENT DEVELOPMENTS IN POWDER DIFFRACTION FILE

INCLUSION OF MODULATED STRUCTURES

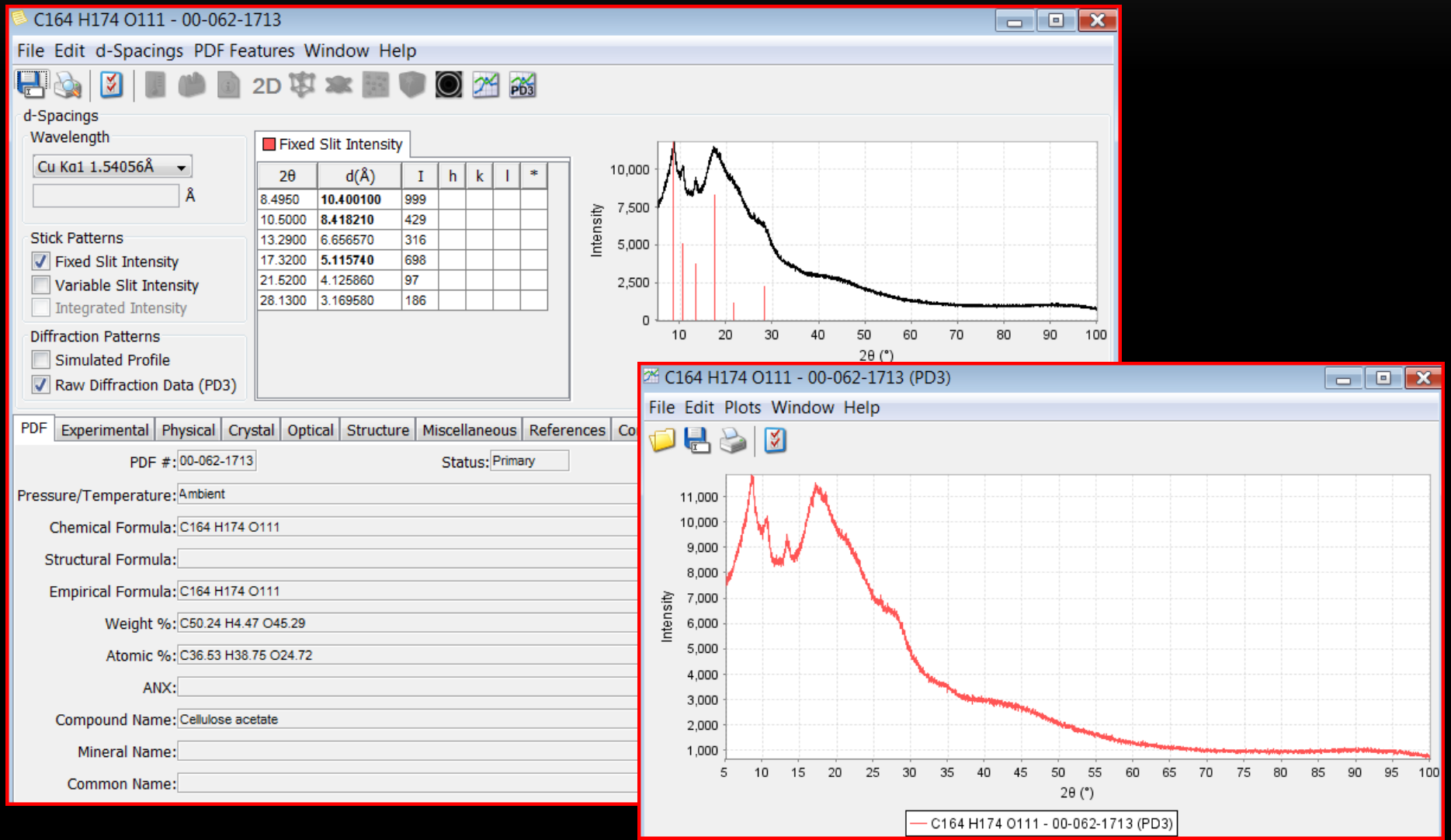


MODULATED STRUCTURES





INCLUSION OF POORLY CRYSTALLINE MATERIALS

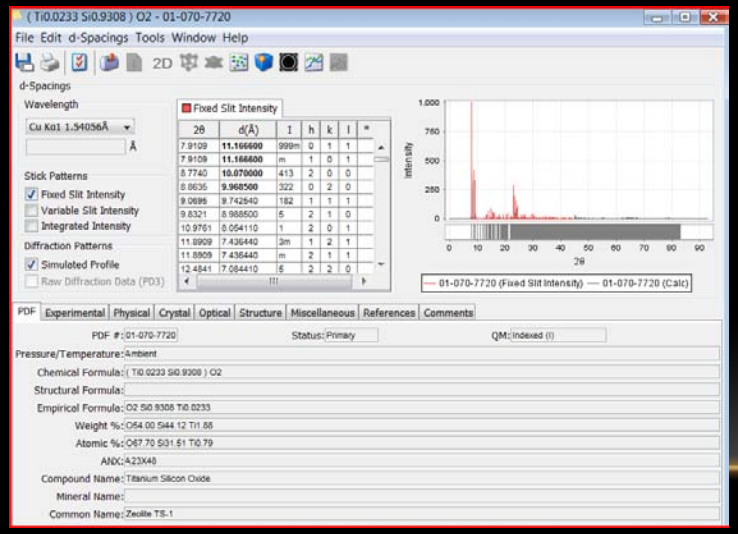


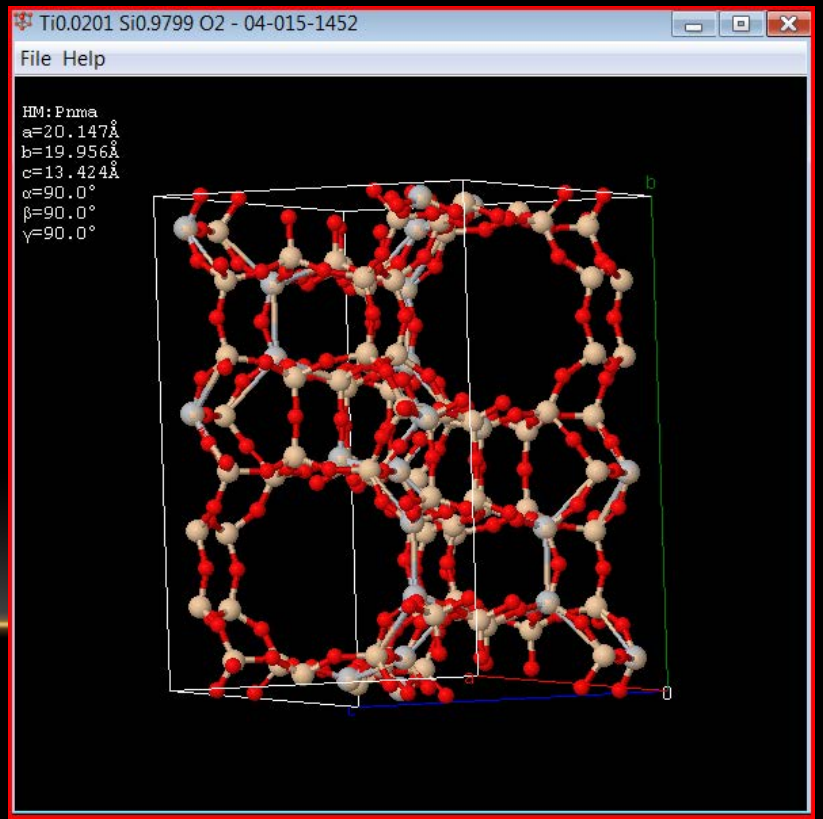
NEW CROSS REFERENCES

- Three new cross references have been introduced to Powder Diffraction File
  - Cross reference to atomic coordinates
  - Cross reference to related phase
  - Grouping of non-ambient temperature data

ATOMIC COORDINATE CROSS REFERENCE

- This cross reference focuses on mapping entries with no atomic coordinates to ones with atomic coordinates
- PDF-4+ 2013 will have 183,193 atomic coordinates out of 340,653 entries
- 36,252 atomic coordinate cross references have been indentified resulting in total number of entries with atomic coordinates to 219,445





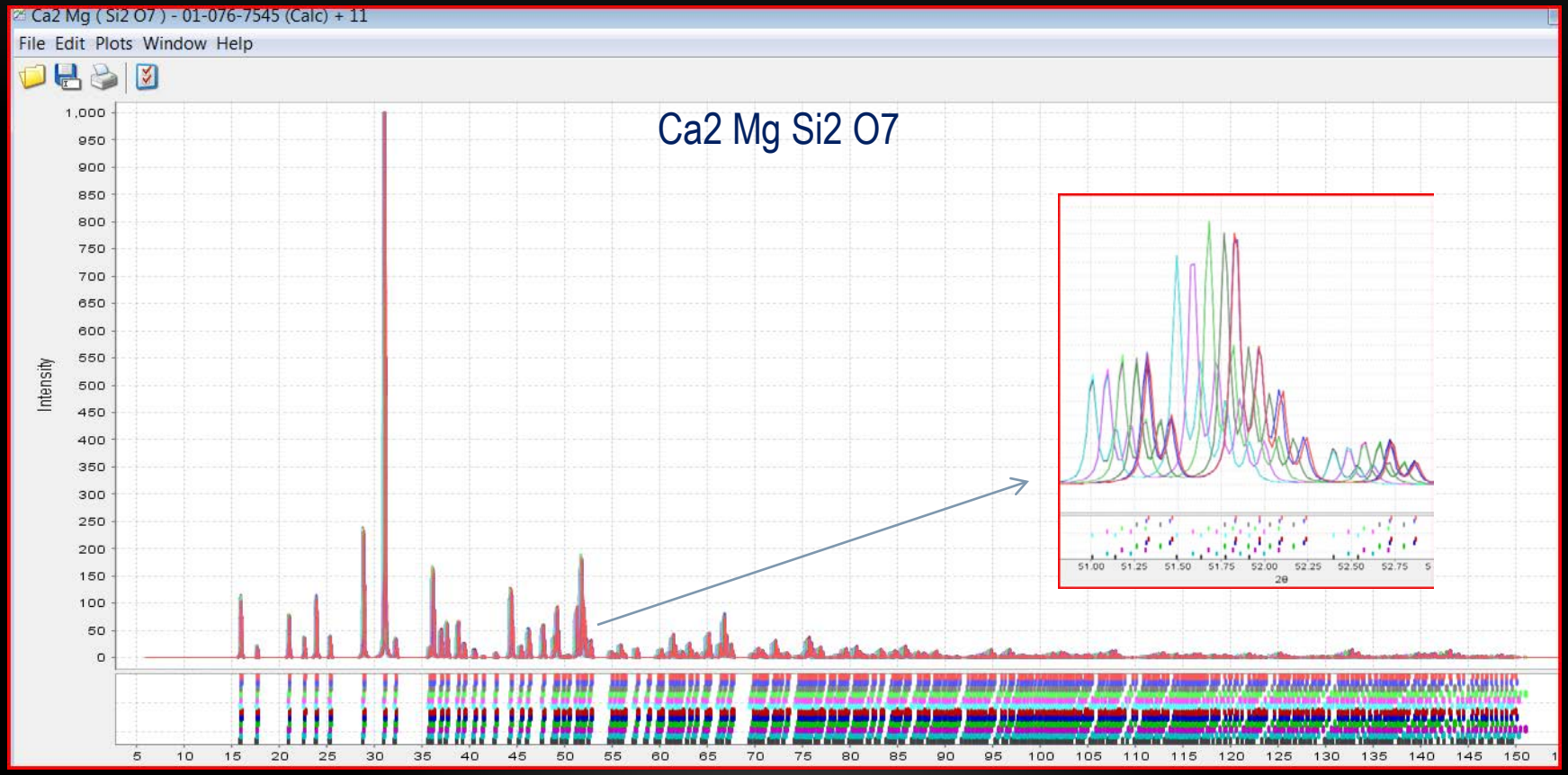
CROSS REFERENCING RELATED PHASES

- When there is no crystal structure available for a given phase, cross referencing has been implemented to relate to the compositionally closest phase
- 13,975 phases with no atomic coordinates have been mapped to the closet composition with atomic coordinates in PDF-4+ 2013

TEMPERATURE SERIES DATA

- For a given chemical composition all the diffraction data collected at various temperatures are grouped together
- Good tool to study coefficient of thermal expansion and phase transition
- Grouping was done for phases that have at least two non-ambient temperature diffraction data
- When temperature is absent, 293 K is assumed and flagged accordingly

NON-AMBIENT TEMPERATURE DATA



There are 4845 distinct temperature series data in PDF-4+ 2013