Coupling the PDF-4/Organics Relational Database to SIeve+, a Hanawalt and Fink Search Indexing Plug-In

J. Faber, J. Blanton and C. A. Weth International Centre for Diffraction Data (ICDD) Newtown Square, PA 19073, USA

Introduction

The International Centre for Diffraction Data (ICDD) has been the reference source for Xray powder diffraction (XRPD) data for over 50 years. The primary information in the PDF is the collection of d-I data pairs, where the d-spacing (d) is determined from the Bragg angle of diffraction, and the peak intensity (I) is obtained experimentally under the best possible conditions for a phase-pure material. These data provide data mining [1capability as well as "fingerprint" 21 identification of the compound because the dspacings are fixed by the geometry of the crystal and the intensities are dependent on the contents of the unit cell. Hence, d-I data may be used for identification of unknown materials by locating matching d-I data in the PDF with the d-I pairs obtained from an unknown specimen.

The PDF has exhibited recent dramatic growth in entry population over the past 5 years and the new PDF-4/Organics results from a collaboration with the Cambridge Crystallographic Data Centre (CCDC). An illustration of organic entry growth for the PDF-4/Organics RDB is given in Table 1.

Identification is the most common use of the PDF, but the presence of considerable supporting information for each entry in the PDF allows further characterization of the specimen. In addition, we wish to demonstrate the use of the PDF-4 relational database as a filter for search-indexing. In this case, we will use a new plug-in for PDF-4 databases, SIeve+, to illustrate search results using Hanawalt and Fink methods.

Category	PDF-4/	PDF-4/	PDF-4/
	Full File	Organic	Organic
	2003	2003	2004
Organic	25,609	147,201	217,077
Compounds			
Inorganic	133,370		3,048
Compounds			
Both Organic	1,931	1,776	1,931
and Inorganic			
Only Inorganic	131,439		1,117
Calc. patterns		122,816	191,468
from CSD			
Drug Activity		4,508	6,343
Pharmaceuticals	2,039	1,192	2,352
Excipients	801	184	1,114
Forensic	3,767	2,015	2,113
Materials			
Pigments	342	284	296
I/Ic	73,087	125,342	195,316
Total Entries	157,048	147,201	218,194

Table 1. Selected entry counts for PDF-4 databases. Please note that because entries can be listed in both the inorganic and organic collection, the total number of distinct entries is obtained from the organic and only inorganic rows in the table.

PDF-4/Organics

The PDF-4 database contains interplanar spacings (d) and relative intensities (I). However, other useful data such as synthesis, physical properties and crystallographic data are also stored in the database. With this new format, we will provide a broader range of analyses, for example, improved quantitative analyses, full pattern display, bibliographic cross referencing, etc. The PDF-4 uses relational database technology that provides pliable access to the database to carry out data mining studies and enhances the pursuit of conventional materials characterization using diffraction techniques (see Faber et al. [2]). In addition to better access to some of the RDB fields, users can also build search criteria by combining individual search conditions using Boolean operators. The availability of logical operators for combining search criteria is very useful for retrieval of relevant information from the database.

Property	Entry Population
Drug Activity	6,343
Pharmaceuticals	2,352
Excipients	1,114
Merck	1,554
Pigments	296
Color	214,501
Density	201,084
Melting Point	64,733
Organic Functional	41,552
Group Designations	
Empirical Formula	All
# of Elements	All
Periodic Table	All
# Searchable Fields	>30

Examples of searchable fields are illustrated in Table 2.

Table 2. Selected searchable fields in the PDF-4/Organics 2004 that can be used as pre-filters for Search-Indexing, using the RDB plug-in, SIeve+.

Search-Indexing using the Plug-in, SIeve+ in the PDF-4/Organics 2004

Manual techniques were first discussed by Hanawalt and these persist for a variety of reasons. The search-indexing plug-in, SIeve+, discussed here follows a traditional path to act as a replacement for paper search manuals published by the ICDD. An advantage to this approach is that Hanawalt [3,4] and Fink methods [5] can be followed in great detail as search-indexing proceeds. Educational benefits accrue following this approach. A more detailed report concerning PCSIWIN, the predecessor of SIeve+ has been given [6].

The Hanawalt search method has been implemented for many years at the ICDD. The method involves sorting the patterns in the PDF according to the d-spacing value of the strongest line. This list is partitioned into discrete d-space intervals defined as Hanawalt groups. A small overlap in d-intervals is employed to reduce the probability of missing powder pattern entries due to uncertainty in the d-space accuracy. Each Hanawalt group

is sorted in order of decreasing d-spacing of the second most intense diffraction line. Subsequent lines are listed in order of decreasing intensity. The analysis rests on the three most intense lines. All reference lines must then be compared to the unknown. The Fink method was designed as an index based on the eight strongest d-spaces in the experimental pattern, but these are ordered in decreasing d-spacing. In short, the Fink method considers the 8 longest of the strongest diffraction lines. To help account for preferred orientation effects, the lines need to be permuted to remove limiting intensity criteria. This permutation process is easily computerized (in contrast to paper manual searches).

Filtering criteria need to be "remembered" while carrying out the search/indexing process. We have developed a "plug-in" for the PDF-4 databases that implements the Hanawalt/Fink strategy, including chemistry, subfile, quality-mark filters and any of the searchable properties in the database (see Table 2).

SIeve+

The basic idea of the plug-in is to provide d,I pairs as input to the program. The d-spacings are in Å and the I's are peak intensity values from the X-ray powder diffraction experiment. The principal input is from an ASCII file that contains the d,I pairs. However, the plug-in can also accommodate 2θ ,I pairs if the first record also contains the wavelength. The uncertainty in d, Δd , can be obtained by taking the derivative of Bragg's Law:

$$\Delta d = d \cdot \cot \theta \cdot \Delta \theta$$
. Eq. 1

In the case of the Hanawalt method, the search window, $SW = \Delta(2\theta_{SW})$ defines the Hanawalt group; the match window, $MW = \Delta(2\theta_{SW})$ defines the criteria used to judge matches for individual lines for each member of the group. The angular

File Se	earch Hel	n i											
110 01		e La sel	1	4 (200	. –	-	1 1 1	1				- 1	
De	· . (†		a 📥 🕯	00	d2 60	H 🗛 🔎	· · · · · · · · · · · · · · · · · · ·	5 🖸 P 💈) E		81	l•	
Curren	t Data R	ecord :	Ex-7-4-1		Num	per of Hit	s : 1977	out of :	218194		(<u>a</u>		
Wavel	length :	1.54056	Г	Chemical Name								8	
nput l	Units: d	-spacing	s	2-(Acetyloxy)-benzoic acid								-	
Searc	h Method	I: Hana	walt	Aspirin	000	000120850							
Search	h Window	• 0 18		Ampicillin (anhydrous)								000291545	
Matak	Mindow	. 0 10		Acetylsalic	ylic acid					020	020600434		
matth	winuow	. 0.10		Acetylsalic	ylic acid					000	010182		
		. EOO		tris(6-Mercaptopurine-N,S)-iron(ii) tetrachloro-iron(iii) chloride 02									
Min. A	ulow GUI	M: 300		tris(6-Merc	captopuri	ne-N,S)-irc	on(ii) tetra	icnioro-iron(iii)	chloride	020	100204		
Min. A Searc	h Line:	M: 500 5.68		tris(6-Mero trans-Diflu	captopurii ioro-(1,4,	ne-N,S)∔rc 8,11-tetra∹	in(ii) tetra azaundeo	icnioro⊣ron(iii) cane) chromiur	n(iii) peri	chlorat 020	639181		
Min. A Searci d1 Ba	h Line:	M: 500 5.68 615.57	45	tris(6-Mero trans-Diflu	captopuri oro-(1,4,	ne-N,S)-irc 8,11-tetra- Botation	azaundeo 1 of 3	icnioro-iron(iii) cane) chromiur I	chioride n(iii) peri	chlorat 020	639181		
Min. A Searci d1 Ra	h Line: nge : 5,	м: 500 5.68 615 - 5.7	45	tris(6-Mero trans-Diflu	captopuri oro-(1,4,	ne-N,S)4rc 8,114etra4 Rotation	n (ii) tetra azaundeo n 1 of 3	ichioro4ron (iii) cane) chromiur I	chioride n(iii) peri	chlorat 020	639181		
Min. A Searci d1 Ra Lines1	nge : 5,	м: 500 5.68 615 - 5.7	45	tris (6-Meri trans-Diflu	captopurii oro-(1,4,	ne-N,S)-irc 8,11-tetra- Rotation	n (ii) tetra azaundei n 1 of 3	ichioro-Iron (III) cane) chromiur I	chioride n(iii) peri	chlorat 020	639181		
Min. A Searci d1 Ra Lines1 D1	nge : 5, D2	5.68 615 - 5.7	45	tris(6-Mero trans-Diflu D5	captopurii oro-(1,4,	ne-N,S)-iro 8,11-tetra- Rotation	n (ii) tetra azaunder n 1 of 3	Ichloro-Iron (III) cane) chromiur	chloride n(iii) peri	chlorat 020	GOM	~	
Min. A Searci d1 Ra Lines1 D1 5.67	h Line: nge : 5, D2 3.929	5.68 5.68 615 - 5.7	45	tris (6-Mero trans-Diflu D5 3.312	captopurii oro-(1,4, D6 4.298	ne-N,S)-iro 8,11-tetra- Rotation D7 3.83	n (ii) tetra azaunder n 1 of 3 D8 2.741	PDF# 020600435	chioride n(iii) peri QM C	chlorat 020	GOM 6192		
Min. A Search d1 Ra Lines1 D1 5.67	D2 11.4	5.68 615 - 5.7 D3 11.374	45 D4 3.286	tris(6-Mero trans-Diflu D5 3.312	captopurii oro-(1, 4, D6 4.298	ne-N,S)-irc 8,11-tetra- Rotation D7 3.83	n (ii) tetra azaunder n 1 of 3 D8 2.741	PDF# 020600435	chioride n(iii) peri QM C	Subfile	GOM 6192 6032		
Min. A Searc d1 Ra Lines1 D1 5.67 5.7	D2 3.929 11.4 11.4	5.68 615 - 5.7 D3 11.374 3.929 3.929	45 D4 3.286 3.23 3.3	D5 3.312 4.3 2.000	D6 4.298 2.743	ne-N,S)-irc 8,11-tetra- Rotation D7 3.83 5.3 5.3	D8 2.741 2.849	PDF# 020600435 000291545	Chioride n(iii) peri QM C S	Subfile	GOM 639181 6192 6032 5436		
Min. A Searci d1 Ra Lines1 D1 5.67 5.67 5.667	D2 3.929 11.4 3.935	5.68 615 - 5.7 D3 11.374 3.929 11.392 11.392	45 D4 3.286 3.23 3.3 3.291	D5 3.312 4.3 3.316 4.3	D6 4.298 3.31 2.743 4.299	D7 3.83 5.3 3.83 5.3 3.836 5.3	D8 2.741 2.849 2.744	PDF# 020600435 000291545 020600434	QM C S	Subfile	GOM 639181 639181 6192 6032 5436 4474		
Min. A Searci d1 Ra Lines1 D1 5.67 5.67 5.67 5.7 5.667 5.7	D2 3.929 11.4 3.935 3.3 2.924	5.68 5.68 615 - 5.7 D3 11.374 3.929 11.392 3.92 11.392 3.92 4.205	45 D4 3.286 3.29 3.3 3.291 11.7 2.202	D5 3.312 4.3 3.316 4.29	D6 4.298 3.31 2.743 4.299 2.15 2.200	D7 3.83 5.3 3.836 2.73 7.649	D8 2.741 2.849 2.744 2.64 2.744 2.6 5.175	PDF# 020600435 0001208501 000291545 020600434 000010182	Chloride n(iii) per- QM C I C I C	Subfile	GOM 639181 639181 639181 6192 6032 5436 4474 4308 2025		
Min. 6 Searcl d1 Ra Lines1 D1 5.67 5.67 5.7 5.667 5.7 5.675 5.675 5.675	D2 3.929 11.4 11.4 3.935 3.3 3.824 5.249	5.68 5.68 615 - 5.7 11.374 3.929 11.392 3.929 11.392 3.92 4.286 4.295	45 3.286 3.29 3.3 3.291 11.7 3.303 2.944	tris(6-Merc trans-Diflu 05 3.312 4.3 4.3 3.316 4.29 4.88 11 266	D6 4.298 3.31 2.743 4.299 2.15 2.928 4.299 2.15 2.928	D7 3.83 5.3 3.83 5.3 3.836 2.73 7.648 4.22	D8 2.741 2.741 2.744 2.744 2.745 2.744 2.744 2.6 5.175 9.752	PDF# 020600435 0001208500 000291545 020600434 020600434 020600434 020785254	QM QM C S C C C C C C	Subfile	639181 639181 6192 6032 5436 4474 4308 3825 2599		
Min. A Searci d1 Ra Lines1 D1 5.67 5.67 5.67 5.667 5.683 5.683	D2 3.929 11.4 11.4 3.935 3.3 3.824 5.185	 5.68 615 - 5.7 11.374 3.929 11.392 3.92 4.285 4.305 7.333 	D4 3.286 3.29 3.3 3.291 11.7 3.303 3.944 3.83	D5 3.312 4.3 3.316 4.29 4.88 11.366 4.49	D6 4.298 3.31 2.743 4.299 2.15 2.928 4.376 4.376 2.491	D7 3.83 5.3 3.836 2.73 7.648 4.234	D8 2.741 2.849 2.744 2.6 5.175 8.752 2.772	PDF# 020600435 000120850 000291545 020600434 020600434 020600434 020785254 020785254 020785254	QM QM C C C C C C C C C	Subfile	639181 639181 6192 5032 5436 4474 4308 3825 3698 3631		
Min. A Searcl d1 Ra Lines1 5.67 5.67 5.67 5.67 5.67 5.675 5.683 5.683 5.683	D2 3.929 11.4 3.935 3.3 3.824 5.185	 5.68 615 - 5.7 D3 11.374 3.92 3.92 4.286 4.305 7.333 4.234 	D4 3.286 3.29 3.3 3.291 11.7 3.303 3.944 3.83	tris (6-Merc trans-Diffu 3.312 4.3 3.316 4.29 4.88 11.366 4.49	D6 4.298 3.31 2.743 4.299 2.15 2.928 4.376 2.491	D7 8,11-tetra= Rotation 3,83 5,3 3,836 2,73 7,648 4,23 4,234	D8 2.741 5.29 2.849 2.744 2.6 5.175 8.752 2.772	PDF# 020600435 00012050 00021545 000021545 000010182 0001018254 000010182 000010182 000010182 000010182 0000010182 0000010182 0000010182 0000010182 00000010182 00000010182 00000010182 00000010182 0000000000	QM C C C C C C C C C C C C C	Subfile 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	GOM 639181 6192 5436 4474 4308 3825 3698 3631		

Figure 1. Hanawalt method applied to an over-the-counter medication. The drug is aspirin. The tablets were ground and standard XRD experiments were performed. A peak-listing program was used to define dspacings and peak intensities for all Bragg lines detected. A wide search window and match window were employed for this analysis.

🖽 Pat	ttern Ma	tching													
File Se	earch Help	5													
De	· • •			00	d2 60		9	JO P	2 0		8 📭	Ex d(A)	Exi	P1 d(A)	P1
Curren	t Data R	ecord :	Ex-7-4-1	1	Num	ber of Hil	ts: 6	out of :	21819	4		11.37	40	11.4	3
Wavel	ength :	1.54056	;	Chemical Name PDF #								5.68	100	5.68	10
Input I	Units : d	-spacing	1s	Aspirin 000120850							120850	5.3	5	5.29	
Searc	Search Method : Hanawalt Zidovudine 000421848								421848	4.88	2	4.89	1		
Search Window : 0.12 Chlortetracycline hydrochloride 000050214								050214	4.3	18	4.3	2			
Match Window : 0.12 Thioridazine hydrochloride 000311931								311931	4.23	5	4.23				
Glipizide 000431753								431753	4.13	2	4.13				
Min. Allow GUM: 500 Glipizide 000451762								451762	3.92	34	3.93	3			
Searc	Search Line: 5.68									3.83	11	3.83			
d1 Range : 5 637 - 5 723 Rotation 1						n 1 of 3	3				3.56	2	3.56		
Liport	1											3.304	29	3.31	2
LINEST	-											3.285	34	3.29	2
D1	D2	D3	D4	D5	D6	D7	D8	PDF#	QM	Subfile	GOM	2.954	3	2.957	1
5.68	11.4	3.93	3.29	4.3	3.31	3.83	5.29	000120850	S	0	5257	2.844	2	2.848	
5.683	9.958	3.979	5.549	6.005	4.137	5.156	4.862	000421848	S	0	1559	2.739	5	2.734	1
5.64	5.24	3.11	7.6	4.06	4.33	3.67	4.23	000050214	B	0	1004	2.604	2	2.6	
5.64	3.55	4.44	6.66	5.14	5.47	4.19	4.96	000311931	В	U	1004	2.49	2	2.488	
5.673 E COA	12.086	8.031	9.764	4.083	0.204	4.939	4.623	000431753	0	0	373	2.456	2	2.458	
3.004	4.005	14.001	3.704	4.031	4.032	13.073	4.333	000431762	10	1.0	020	2.222	2	2.226	
												2.15	2		
					1000							2.143	2	2.145	
1000	688	264	423	1000	343	1000	539								Γ
P1 00	00120850) Aspiri	n					Int. Sca	le Fac	tor: 1.017	100%				

dependence of the search window and match window are defined by Eq. 1. The SIeve+ results illustrated in Figures 1 and 2 show that aspirin is the unknown phase.

In Figure 1, wide search and match windows were chosen and the result is a Hanawalt group that contains 1,977 entries. Each of these reference patterns are then compared line by line against the "unknown".

In contrast, Figure 2 shows the results of coupling the PDF-4/Organics RDB as a pre-filter to

the search indexing functions in SIeve+. Only 6 reference patterns need be compared to the unknown phase. It important to recognize that the Pharmaceutical subfile search in the PDF-4 was executed and the search results were conveniently transmitted to SIeve+.

Summary

We have outlined how the PDF-4/Organics can be used as a pre-filter to search-indexing using classical Hanawalt and Fink methods. The advantage of pre-filtering centers on the ability to use any combination of the searchable properties in the RDB. These can be applied to searchindexing using SIeve+. There were no filtering protocols available with older search-index paper

Figure 2. Hanawalt method applied after prefiltering using the PDF-4/Organics 2004 RDB. Notice that the Hanawalt group has been reduced to just 6 candidate reference materials. manuals. In contrast for the plug-in, SIeve+, the effort required to examine large numbers of entries in the Hanawalt or Fink group is minimized and relevant solutions to search-indexing are illustrated. This streamlines the search-indexing process.

References

[1] Faber, J., Kabekkodu, S.N. and Jenkins, R. (2001), International Conference on Materials for Advanced Technologies, Singapore, unpublished; Kabekkodu, S.N., Faber, J. and Fawcett, T., Acta Cryst., Vol. B58, 333-337 (2002). [2] Faber, J. and Fawcett, T., Acta Cryst., Vol. B58, 325-332 (2002). [3] Hanawalt, J. D. and Rinn, H. W., Ind. Eng. Chem. Anal., 8, 244 (1936); Hanawalt, J. D., Advances in X-Ray Analysis, 20, 63-73 (1976). [4] Hanawalt, J. D., Cryst. in North America, Apparatus and Methods, American Crystallographic Association, Chapter 2, 1983, 215-219. [5] Bigelow, W. and Smith, J.V., ASTM Spec. Tech. Publ. STP 372, 54-89 (1965). [6] Faber, J., Weth, C. A. and Jenkins, R. (2001), Materials Science Forum, Vol. 378-381, 106-111 (2001); Faber, J. and Weth, C. A., J. Powder Diffraction, Vol. 19, 26-30 (2004).