**Search/Match of Pharmaceutical Formulations**

These formulation data sets are designed to demonstrate phase identification methods in formulation analyses. The results obtained were achieved using the PDF-4/Organics database and the embedded identification software used with this database. These data sets have been tested with several commercial search/match software systems and a range of databases. If you use other databases or software you will not obtain the full correct results. Most databases, not published by ICDD, are based on crystalline materials and do not have appropriate references for non-crystalline materials. The ICDD has many proprietary powder pattern references, both crystalline and non-crystalline, developed through their research programs and grants that are not available in other databases. If you use a large database of crystalline materials, commercial search/match software will result in false positive results. With large databases the software will try to match the peaks, and will find candidates, even if the appropriate references are not available.

Technique is always important to obtain good results. There are low concentrations phases in several of these experimental patterns. One has to be sure that all the peaks are identified, which often means that highly automated methods should not be used. Remember that some of these patterns have amorphous or nanocrystalline phases so a polynomial background fit is inappropriate as it will remove intensity from non-crystalline or nano phases of interest.

***In order of difficulty -***

**Celebrex®** *–* This analysis isrelatively easy, the main API and excipient should be 1 and 2 on the search/match candidate list*.* The API is celecoxib (PDF 02-075-7351) and the excipient is highly crystalline alpha lactose monohydrate (PDF 02-063-2272). The remaining unidentified peaks at low angle are due to magnesium stearate dihydrate.

**Centrum Performance® –** Finding a few phases is easy, finding all the phases is difficult unless both good software and technique is used.This particular data set was taken on a laboratory diffractometer equipped with an incident beam Ge monochromator, resulting in copper alpha one incident radiation. This is important since the data are well resolved and one should find many phases. This data set has been used many times at various ICDD educational classes. Most people will identify 5-6 phases, some will get 8, and a few will get 9-12. Vitamin C, KCl, ZnO, Brushite, Monetite and Calcite should be readily identified.

To get more phases you have to pay attention to several doublets at low angles since the peaks in the doublets are from different phases, and several weak intensity peaks are the key to finding other vitamin phases. If you do this, you should find iron fumarate, niacin, magnesium oxide, manganese sulfate monohydrate. Some people will find riboflavin, but usually after careful elimination of all other phases and an analysis of residual peaks, since only the largest peaks of this phase are visible. Careful examination of any remaining residual will show broad features around 22 degrees, microcrystalline cellulose Iβ.

This data set has been tested on several commercial automated search match systems. They typically detect the six major phases, primarily because they miss low intensity peaks of the minor phases with automated background and peak identification algorithms. If the user manually optimizes the background subtraction and peak finding algorithms, then minor phases will be identified with most commercial software packages.

**Motrin®** – The API acetaminophen **(**PDF 02-076-2281) is the main phase, easy to detect. Additional phases include cellulose and anatase. If you are using total pattern fitting techniques you will need to add the PD3 pattern of amorphous cellulose (PDF 00-060-1501) to match the experimental data.

**Benicar®** –Phase identification of this data set is moderately difficult. The tablet formulation consists of the excipient, alpha lactose monohydrate, in high concentration with an API at low concentration (9.4%). The challenge in the identification of alpha lactose monohydrate is that a small crystallite size causes the major peaks to broaden and merge. The resulting shifted peak positions and deconvoluted intensities will make this phase more difficult to identify with most automated software packages. The lactose monohydrate needs to be identified to successfully identify the residual API peaks of olmesartan medoximol (PDF 00-067-1352). The API phase is also slightly oriented. This is where a pattern fitting method excels since you can see the changes if you apply an orientation model to appropriately match all the peak intensities. If everything is done correctly you should be able to not only determine the phases but get a good quantitative phase match.

**Azor®** – This is a challenge but not impossible. There is a transparency error in this data set. Without a transparency correction the API has an average displacement of 0.14° 2θ, which puts this about 50th on the candidate list. The API is olmesartan medoximol(PDF 02-069-6187) and the main excipient is nano crystalline cellulose Iβ (PDF 00-060-1502). The analyst has to deeply search the candidate list because the crystalline API will have a low score due to the shift. If the shift is recognized and corrected the score will dramatically increase and make it easier to identify the cellulose Iβ. Finally, if both of these phases are accounted for one can identify the second API, amlopidine besylate (PDF 00-060-1141), in the residual pattern. (Note: The score refers to the goodness of match algorithm used in the software to match references to the experimental data. The specific algorithm will vary from software to software, but all commercial systems use a scoring algorithm. In some software a low value is the best match and in others a high value is the best match, so you would need the help file to see how the scoring works in your particular software.)

This data set is also an example where the use of a similarity index can help identify the major phases. Using the Integral Index similarity algorithm one can identify both cellulose and olmesartan medoximol by their pattern profiles. The user would notice the broad peaks and compare the experimental data to PD3 patterns of excipients, identifying cellulose Iβ. Similarly, a search of all API’s would identify olmesartan medoximol. In these cases, the major phases are a large portion of the scattering intensity, resulting in successful identification. In both cases the user would see a visible shift in the display comparison of reference to the experimental data, signaling a correction (transparency). Integral index calculations use point by point comparisons on data sets with thousands of points, so the use of subfiles speeds the analyses. The matches show up in the top candidates prior to transparency correction, but not the top candidate until after the correction is made. The user would have to check the visible matches of the top dozen candidates from the uncorrected raw data and the match is visibly obvious since you can easily observe the shift as all peaks are shifted in the same direction.

The analysis and correction of the transparency error is critical to find minor phases independent of the use of conventional or similarity index methods. The GOM, which is a goodness of fit for the reference pattern, of the olmesartan medoximol is 2600 in the original data set. The GOM increases to 5200 when the correction is applied to the experimental data and the references goes to the top of the candidate list. Since the experimental data are corrected, the identification of all other phases proceeds accordingly. To get a perfect match with total pattern fitting methods, the user would have to add a little amorphous cellulose (PDF 00-060-1502) and identify the magnesium stearate. Once scaled these five phases will match the experimental data perfectly.

It should be noted that some search/match software will not identify phases if the data sets have large displacement or transparency errors. The user might have to check the search windows used in the search/match software, this is often a user-defined parameter and the windows can often be enlarged (check the help files or user manual). In ICDD’s PDF-4/Organics software this transparency error was within the default limit, so the API showed on the candidate list.

.