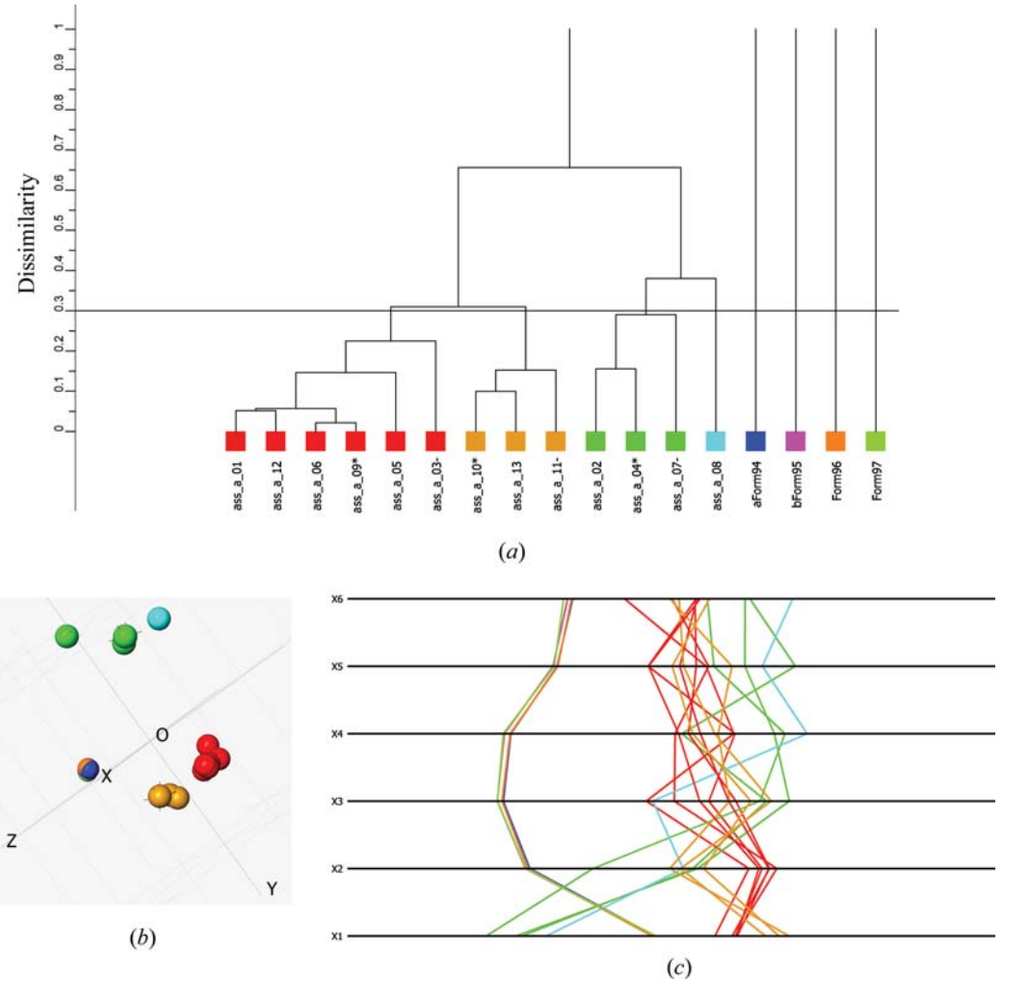

Figure 3.8.10
The aspirin data including data from five amorphous samples. (*a*) The resulting dendrogram and (*b*) the corresponding MMDS plot. (*c*) The parallel-coordinates plot.

$a = 7.14$, $b = 7.65$, $c = 5.83$ Å with $Z = 4$; between 357 and 398 K it crystallizes in the tetragonal space group $P\bar{4}2_1m$ with $a = 5.719$, $c = 4.932$ Å, $Z = 2$, and above 398 K it transforms to the cubic space group $Pm\bar{3}m$ with $a = 4.40$ Å and $Z = 1$. PXRD data containing 75 powder patterns taken at intervals of 3 K starting at 203 K using a D5000 Siemens diffractometer and Cu $K\alpha$ radiation with a 2θ range of 10–100° were used (Herrmann & Engel, 1997). Fig. 3.8.11(*a*) shows the data in the 2θ range 17–45°.

The visualization of these data following cluster analysis is shown in Fig. 3.8.11(*b*) using an MMDS plot on which has been superimposed a line showing the route followed by the temperature increments. The purple line follows the transition from a mixture of forms IV and V at low temperature (red) through form IV (yellow), form II (blue) and finally form I at high temperature (green). This is an elegant and concise representation of the data in a single diagram.

3.8.7. Quantitative analysis with high-throughput PXRD data without Rietveld refinement

Since mixtures are so common in high-throughput experiments, and indeed in many situations with multiple data sets, it is useful to have a method of automatic quantitative analysis. The quality of data that results from high-throughput crystallography makes it unlikely that an accuracy better than 5–10% can be achieved but, nonetheless, the identification of mixtures can be carried out by whole-profile matching. First a database of N pure phases is created, or, if that is not possible, then the most representative patterns with appropriate safeguards can be used. Assume that there is a sample pattern, S, which is considered to be a mixture of up to N components. S comprises m data points, $S_1, S_2, \ldots, S_m$. The N patterns can be considered to make up fractions $p_1, p_2, p_3, \ldots, p_N$ of the sample pattern. The best possible combination of the database patterns to fit the sample pattern is required. A system of linear equations can be constructed in which x_{11} is measurement point 1 of pattern 1 *etc.*:

$$\begin{aligned} x_{11}p_1 + x_{12}p_2 + x_{13}p_3 + \ldots + x_{1N}p_N &= S_1, \\ x_{21}p_1 + x_{22}p_2 + x_{23}p_3 + \ldots + x_{2N}p_N &= S_2, \\ &\vdots \\ x_{m1}p_1 + x_{m2}p_2 + x_{m3}p_3 + \ldots + x_{mN}p_N &= S_m. \end{aligned} \tag{3.8.26}$$

Writing these in matrix form, we get

$$\begin{bmatrix} x_{11} & x_{12} & x_{13} & \cdots & x_{1N} \\ x_{21} & x_{22} & x_{23} & \cdots & x_{2N} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ x_{m1} & x_{m2} & x_{m3} & \cdots & x_{mN} \end{bmatrix} \begin{bmatrix} p_1 \\ p_2 \\ \vdots \\ p_N \end{bmatrix} = \begin{bmatrix} S_1 \\ S_2 \\ \vdots \\ S_N \end{bmatrix} \tag{3.8.27}$$

or

$$\mathbf{xp} = \mathbf{S}. \tag{3.8.28}$$

A solution for S that minimizes