Application of Multidimensional Scaling to Preformulation Sciences: A Discriminatory Tool to Group Microcrystalline Celluloses

Celine Valeria LIEW,[a] Josephine Lay Peng SOH,[a] Fei CHEN,[b] Daming SHI,[b] and Paul Wan Sia HENG*,[a]

[a] *Department of Pharmacy, National University of Singapore; S117543, Singapore: and* [b] *School of Computer Engineering, Nanyang Technological University; S639798, Singapore.* Received December 11, 2004; accepted July 9, 2005

Pre-formulation studies constitute the first step of any pharmaceutical product development and manufacture. Establishment of a comprehensive library of critical physical, chemical, biological and mechanical properties of all materials used for a formulation can be costly, tedious and time consuming, despite its importance in quality manufacturing management. This study seeks to demonstrate the pharmaceutical application of multidimensional scaling (MDS) by incorporating it as a pre-formulation tool for grouping an expanded range of microcrystalline celluloses (MCC). MDS presents the various MCC grades in two-dimensional space based on their torque rheological properties; thus conferring an extra dimension to the pre-formulation tool to facilitate the visualization of the relative positions of each MCC grade. Through this work, the utility of MDS for expediting pre-formulation studies, in particular, grouping of excipients that are available in different brands and grades can be amply exemplified.

Key words pre-formulation; multidimensional scaling (MDS); microcrystalline cellulose (MCC); rheological property

Pre-formulation studies constitute the first phase of any product development and involve a battery of tests to determine the physical, chemical, mechanical and biological characteristics of a substance. A comprehensive library of this pertinent information provides a valuable resource in the subsequent optimization and processing steps which in turn, ensures product quality. However, the extensive volume of characterization tests and the wide range of expensive equipment required for these tests make it tedious and costly to perform pre-formulation tests.

Pharmaceutical processes are often dictated by complex interactions between the formulation, processing and environmental factors. As these changes occur simultaneously, it is very difficult to quantify or determine the effects of individual factors using one-at-a-time approaches. In view of the difficulties and shortcomings associated with conventional methods of material characterization, pre-formulation tools are developed. In particular, tools that can evaluate the effects of the individual components or serve as performance predictors are extremely valuable.

Modern computational techniques have been widely utilized in the fields of neuroscience, computing, socio-economic studies and psychotherapeutics. On the contrary, their uses in the pharmaceutical sciences have been few and far between. It was only in recent years that computational methods were more increasingly applied in pharmaceutical research. One of the most popular approach involves artificial neural networks (ANN) which is a training system employed to recognize patterns in data sets. In pre-formulation studies, ANN has been used to determine and predict the physicochemical properties of amorphous polymers.[1] When used in combination with data clustering, ANN was able to achieve an objective grouping of several grades of microcrystalline celluloses (MCC) and identify the most important physical property governing their rheological behaviors.[2] With this grouping, there is greater flexibility for manufacturers when it comes to the choice of MCC, one of the most indispensable excipient in extrusion-spheronisation.

Multidimensional scaling (MDS) is a technique that involves the analysis of similarity or dissimilarity between a given set of factors or objects.[3] Every factor will be represented as a point in the geometric space and the distance between points are indicative of the differences between them. In short, it provides a visual representation of the distribution of various factors or parameters in a two-dimensional space. Although MDS was initially developed for the field of computational and statistical studies, it was subsequently employed in diverse disciplines including socio-epidemiology, psychotherapy and biological sciences to analyze patterns and trends that would otherwise be too time-consuming or difficult to perform using existing methods.

Despite its advantages and benefits, pharmaceutical applications of MDS have been lacking. In fact, it was only reported by Barr *et al.*[4] recently as a means to analyze and sort X-ray diffractograms into related clusters which can be further characterized to identify the polymorphs present. Other approaches for matching and sorting X-ray diffractograms include the use of specialized computer softwares which employ both parametric and non-parametric statistics. Compared to these methods, MDS was able to overcome problems that arose from using peaks or representative peaks in the diffractogram.[4] This allows a more accurate way of clustering and matching of unknown diffractograms with the existing ones in the database.

The above work illustrates the usefulness of MDS. It also highlights the importance of harnessing the potential of this technique to expedite and facilitate many processes in the pharmaceutical industry. This current work seeks to demonstrate the applicability of MDS as a pre-formulation tool in the grouping of pharmaceutical excipients, in particular, MCC which is available in many grades from different sources using their torque rheological profiles. Rheological profiles of wetted masses containing MCC have been reported to be related to the performance of MCC in extrusion-spheronisation[5—7] and were thus chosen for data modeling and clustering in this current study. Additionally, the influence of physicochemical properties of MCC on the rheological properties of their wetted masses has been reported in earlier studies.[2]

* To whom correspondence should be addressed. e-mail: phapaulh@nus.edu.sg

Table 1. Physical Properties[a] of Various MCCs

MCC	$\bar{X}$ (μm)	X_{cr} (%)	$V_{low\ P}$[b] (ml g^{-1})	$V_{high\ P}$[b] (ml g^{-1})	V_{total}[b] (ml g^{-1})	ε[c] (%)	ρ_b (ml g^{-1})	ρ_t (ml g^{-1})	$W_{710\,\mu m}$[d] (% w/w)	W_s[d] (% w/w)
Ceolus KG 801	66.65	78.61	2.04	0.57	2.61	76.95	0.19	0.30	37.17	0.064
	(0.39)	(8.89)	(0.05)	(0.02)	(0.07)	(2.55)	(0.00)	(0.00)	(0.75)	(0.003)
Ceolus KG 802	73.19	80.23	1.89	0.28	2.19	69.56	0.23	0.36	37.06	0.056
	(0.85)	(4.34)	(0.02)	(0.12)	(0.12)	(1.35)	(0.01)	(0.01)	(1.29)	(0.016)
Viva Pur 101	68.04	73.52	1.22	0.30	1.52	65.35	0.32	0.45	36.53	0.065
	(0.25)	(3.23)	(0.07)	(0.01)	(0.08)	(1.93)	(0.00)	(0.00)	(0.83)	(0.003)
Pharmacel 101	62.24	72.37	1.11	0.49	1.60	62.29	0.30	0.44	37.64	0.062
	(0.22)	(2.53)	(0.10)	(0.02)	(0.11)	(0.86)	(0.00)	(0.00)	(0.92)	(0.001)
Pharmacel 102	143.44	73.08	1.22	0.23	1.52	61.55	0.34	0.44	37.50	0.065
	(0.66)	(4.14)	(0.04)	(0.04)	(0.07)	(3.60)	(0.00)	(0.00)	(0.62)	(0.001)
Comprecel M101	77.53	75.76	1.33	0.40	1.73	66.11	0.30	0.48	37.10	0.067
	(1.25)	(1.97)	(0.01)	(0.00)	(0.01)	(2.34)	(0.00)	(0.00)	(1.09)	(0.004)
Comprecel M102	118.68	74.42	1.30	0.38	1.67	66.38	0.34	0.50	34.98	0.059
	(1.45)	(2.12)	(0.02)	(0.02)	(0.05)	(2.15)	(0.01)	(0.01)	(0.86)	(0.007)

a) Standard deviations are provided in parentheses. *b*) $V_{low\ P}$: specific cumulative intruded mercury volume into the pores at low pressures (0 to 0.172 MPa); $V_{high\ P}$: specific cumulative intruded mercury volume into intraparticulate pores (10 to 0.006 μm) at high pressures (0.172 to 207 MPa); V_{total}: total specific intrusion volume from 0 to 207 MPa, calculated as the sum of $V_{high\ P}$ and $V_{low\ P}$. ε: percent porosity. *c*) W_s: spheronisation water sensitivity of the MCCs. It measures the tolerance of MCC to added moisture. $W_{710\,\mu m}$: predicted water requirements for producing pellets of mean size 710 μm.

Experimental

Materials A total of 14 MCC grades were evaluated in this study, including 3 additional MCC grades that were added to the 11 MCCs used previously.[2] These 3 additional MCCs were Comprecel M101, Comprecel M102 (Ming Tai Chemical Co. Ltd., Taiwan, Republic of China) and Ceolus KG 802 (Asahi Kasei Chemical Corporation, Osaka, Japan). The other 11 MCC grades were: Avicel PH 101, PH 102, PH 301, PH 302, Ceolus KG 801 from Asahi (Japan), Celex 101 (ISP, Wayne, NJ, U.S.A.), Emcocel 50 M, Prosolv 50 M (Mendell, Patterson, NJ, U.S.A.), Viva Pur 101 (J. Rettenmaier & Sohne, Holzmulle, Germany), Pharmacel 101 and Pharmacel 102 (DMV, Veghel, The Netherlands). Distilled water was used as the granulating liquid.

Methods. Physical Characterization Particle size ($\bar{X}$), bulk (ρ_b) and tapped (ρ_t) densities, % crystallinity (X_{cr}), micromeritic parameters ($V_{low\ P}$, $V_{high\ P}$, V_{total} and ε) and extrusion-spheronisation parameters (W_s and $W_{710\,\mu m}$) were determined according to the methods described by Heng and Koo.[8]

Physical properties of 11 MCC grades (Avicel PH 101, PH 102, PH 301, PH 302, Ceolus KG 801, Celex 101, Emcocel 50 M, Prosolv 50 M, Viva Pur 101, Pharmacel 101 and Pharmacel 102) were previously determined and reproduced here in part for ease of comparison. They were presented along with those of the 3 MCC grades determined in this current study (Table 1).

Rheological Profiles of MCC Grades Rheological profiles of the 3 additional MCCs were determined according to the methods described by Soh *et al.*[2] Fifteen grams of MCC powder were added into the mixer bowl of the mixer torque rheometer (MTR, Caleva Process Solutions, Dorset, U.K.) and mixed for 30 s. The mean torque generated by the dry powder was recorded. Water amounting to 75, 90, 100, 110, 125, 130, 150, 150, 175, 180, 190, 200 and 220% w/w of MCC powder were added as a bolus dose to the dry powder. All the water was evenly distributed throughout the dry powder while the blades turned. The total mixing time was 46 min for all the MCCs and respective torque values were logged at regular intervals. Triplicates were performed for each MCC at all the mixing time and water additions tested. Their torque rheological profiles were presented in Fig. 1.

Data Classification Using ANN A radial basis function (RBF) network with 20 hidden neurons (Fig. 2) was employed to model all the MCCs where the input layer consists of mixing time and water : MCC ratio (x_1, x_2) and the output is torque value. Compared to other network topologies, the RBF network was found to have better scaling properties.[9] A detailed description of the RBF network used for data clustering was reported previously.[2] For our given data set of 260 data points, 20 hidden layers were deemed to be sufficient for data modeling.

Sampling data of 14 MCCs were fed into the ANN to test the mean square errors (MSE). The MSE values range from 0 to 1 and (Table 2) denoted the degree of dissimilarity between any pair of MCC tested. They were used in the subsequent data clustering step that was performed to bring about the grouping. An MSE value of 0 indicates that the MCCs are exactly the same. All the parameters used for the ANN and data clustering have been detailed by Soh *et al.*[2] Briefly, the membership value of every input vector (MCC) will be calculated based on the transfer function stipulated by the RBF network.[2]

The input threshold value (IT) refers to the minimum membership value to decide if an input vector (MCC) can be assigned into any existing group. A given MCC grade will be assigned into an existing group if its membership value is lower than the IT value and *vice versa*. The IT value ranges from 0 to 1. The stringency of this grouping criterion can be regulated by adjusting the IT values. When IT = 1, the MCC grades must be exactly the same before they can be clustered together, therefore all the 14 MCCs will end up individually in 14 different groups at this IT value. Conversely, IT = 0 meant that even MCC grades which are entirely different will still be grouped together. In this case, the MCC grades would not be differentiable. Hence, these 2 extreme values are not useful or practical in achieving an objective grouping of the MCC grades. In view of these considerations, the IT value for Level One stringency (more stringent) was set at 0.9. It was reduced to 0.8 for Level Two.

Multidimensional Scaling (MDS) MDS begins with a distance matrix $\mathbf{R}$ of K data vectors $\{\mathbf{X}_1, \ldots, \mathbf{X}_K\}$ and generates a set of corresponding vectors $\{\mathbf{Y}_1, \ldots, \mathbf{Y}_K\}$ in a 2D or 3D space.[10,11] In the projection procedure, the difference between the two distance matrices is measured by a distortion function $E(\mathbf{r})$. The relative positions among vectors will be preserved by minimizing the error function. MDS algorithm can be implemented in the following steps:

Step 1: Distance of data vectors in the feature space is calculated using the following equation.

$$R_{ij} = \|\mathbf{X}_i - \mathbf{X}_j\| \tag{1}$$

Step 2: A dissimilarity (similarity) matrix is constructed as follows using the MSE values from the RBF network.

$$\mathbf{R} = \begin{bmatrix} 0 & 0 & \cdots & 0 \\ R_{21} & \ddots & \cdots & 0 \\ \vdots & \cdots & \ddots & \vdots \\ R_{K1} & \cdots & R_{K(K-1)} & 0 \end{bmatrix} \tag{2}$$

Step 3: A non-negative function known as distortion function, $E(\mathbf{r})$, is defined and minimized according to Eq. 3.

$$E(\mathbf{r}) = \sum_{i>j}^{K} (R_{ij} - r_{ij})^2 \tag{3}$$

where $r_{ij} = \|\mathbf{Y}_i - \mathbf{Y}_j\|$ is the distance in the target space, $\mathbf{Y} \in \mathbf{R}^2$ or $\mathbf{R}^3$ and $\mathbf{R}$ is a real number.

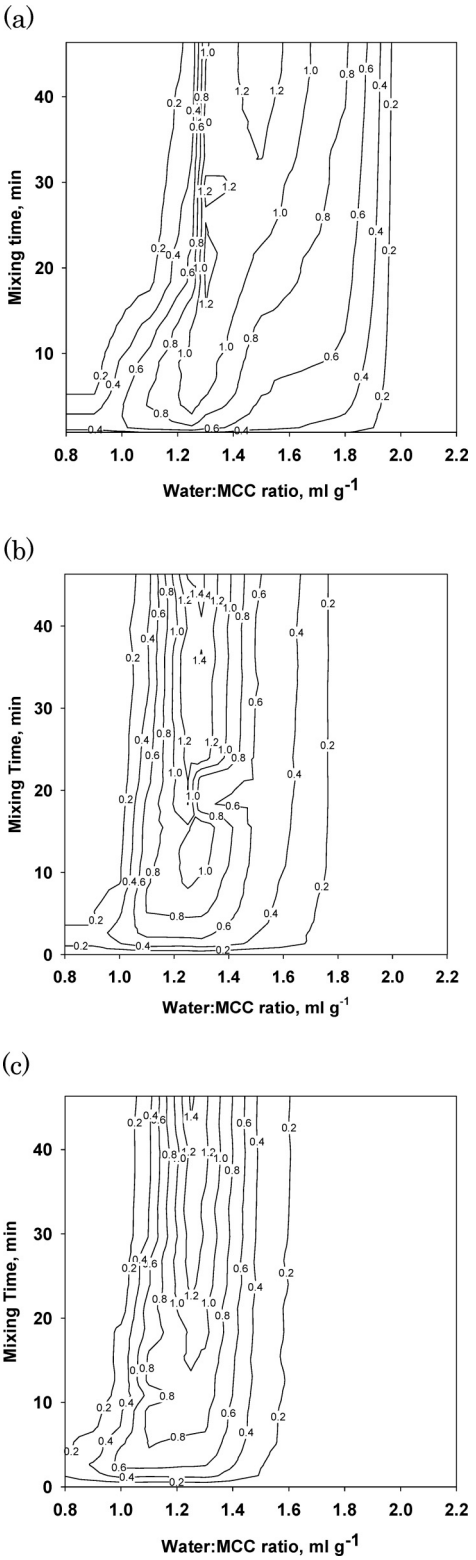

Fig. 1. Variation in Measured Torque [N m] with Mixing Time and Water : MCC Ratio [ml g^{-1}] for (a) Ceolus KG 802, (b) M101, (c) M102

Results and Discussion

Visual Comparison of the MCC Grades Using MDS Using the MDS visualization algorithm, the clustering results in 2D space, with and without PH 301 and PH 302, were presented in Fig. 3. Clearly, the 2 high density MCCs, PH 301 and PH 302, were distinctly different from the other MCCs

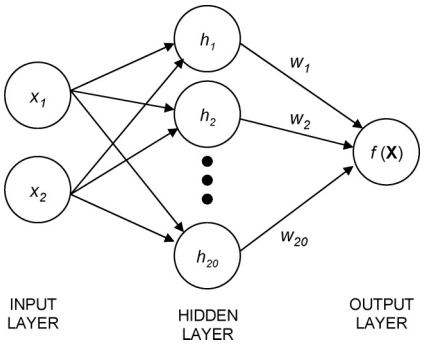

Fig. 2. Architecture of the RBF Network

(Fig. 3a). Hence, compared to them, the other 12 grades of MCC seemed to be alike as shown by their extensive overlapping in the 2D space (Fig. 3a). However, the differences between the remaining 12 MCC grades were better illustrated when these 2 MCCs were excluded from the same 2D space (Fig. 3b).

Figure 3b shows the high degree of similarity between Comprecel M101, Comprecel M102, Pharmacel 101 and Pharmacel 102 and Viva Pur. Similarly, Emcocel and Prosolv are grouped together based on their close relative proximity. On the other hand, Ceolus KG 802 was comparatively further apart from PH 102 and Ceolus KG 801, thus they were initially placed in separate groups under Level One stringency as shown in the following section (Table 3). A relaxation in the stringency of the clustering parameters to Level Two resulted in the assignment of Ceolus KG 802 into the group with the 2 abovementioned MCCs, as well as Celex. This observation reflected the allowance conferred to the stringency of the clustering parameters. In the case of PH 101, it was deemed to be too far apart from the other MCCs to be assigned into any existing groups.

Grouping of MCC Grades Using ANN and Data Clustering The 3 additional MCC grades were assigned into 2 of the existing groups after their rheological data were added to the previous data set used in the training of the neural network and data clustering (Table 3).

Under Level One, Ceolus KG 802 was grouped separately from Ceolus KG 801. When the stringency level was reduced to Level Two, they were grouped together. This observation indicated that these 2 MCC grades were not exactly equivalent in terms of their interaction with water, despite their supposed close similarity. Close inspection of their physical properties (Table 1) revealed that Ceolus KG 802 was of a larger mean particle size and denser than Ceolus KG 801 (ANOVA, $p<0.05$) although both MCCs can be regarded as the more porous grades among the 14 MCCs studied. Furthermore, the slightly lower $W_{710\,\mu m}$ and W_s values of Ceolus KG 802 (ANOVA, $p<0.05$) indicated that less water was required to produce spheroids of 710 μm and it was comparatively less sensitive to changes in water content. This decreased sensitivity was also reflected in the broader contour lines on its contour plot, especially at higher water : MCC ratios (Fig. 1) which meant that the wetted mass was able to maintain the same consistency over a wider range of water : MCC ratios.

Physical characteristics of M101 and M102 did not reveal many notable differences except in their mean particle size,

Table 2. Calculated MSE by Modeling the Torque Measurements of MCCs Using RBF

Testing MCC	Training MCC											
	Avicel PH 101	Avicel PH 102	Avicel PH 301	Avicel PH 302	Ceolus KG 801	Ceolus KG 802	Celex 101	Emcocel 50 M	Prosolv 50 M	Viva Pur 101	Pharmacel 101	Pharmacel 102	Comprecel M101	Comprecel M102
Avicel PH 101	0	0.123	0.908	0.875	0.127	0.065	0.096	0.061	0.076	0.196	0.188	0.180	0.188	0.216
Avicel PH 102	0.130	0	0.809	0.711	0.014	0.068	0.046	0.048	0.040	0.032	0.033	0.033	0.037	0.045
Avicel PH 301	0.676	0.576	0	0.147	0.615	0.656	0.709	0.536	0.556	0.514	0.544	0.581	0.565	0.518
Avicel PH 302	0.663	0.510	0.142	0	0.534	0.622	0.648	0.509	0.531	0.441	0.460	0.488	0.466	0.414
Ceolus KG 801	0.128	0.017	0.873	0.751	0	0.049	0.035	0.053	0.064	0.049	0.043	0.034	0.037	0.057
Ceolus KG 802	0.065	0.071	0.877	0.811	0.048	0	0.034	0.053	0.082	0.128	0.116	0.107	0.111	0.150
Celex 101	0.098	0.046	0.963	0.870	0.033	0.033	0	0.080	0.086	0.097	0.078	0.064	0.084	0.120
Emcocel 50 M	0.061	0.044	0.746	0.698	0.052	0.054	0.081	0	0.025	0.090	0.084	0.095	0.092	0.095
Prosolv 50 M	0.075	0.046	0.803	0.764	0.064	0.083	0.088	0.024	0	0.076	0.083	0.096	0.083	0.091
Viva Pur 101	0.197	0.039	0.739	0.629	0.051	0.132	0.101	0.089	0.077	0	0.014	0.021	0.020	0.015
Pharmacel 101	0.192	0.037	0.782	0.661	0.043	0.117	0.080	0.087	0.080	0.011	0	0.014	0.018	0.017
Pharmacel 102	0.189	0.034	0.836	0.696	0.031	0.104	0.065	0.097	0.090	0.015	0.012	0	0.015	0.025
Comprecel M101	0.180	0.043	0.809	0.658	0.037	0.115	0.084	0.088	0.083	0.018	0.020	0.019	0	0.013
Comprecel M102	0.217	0.052	0.743	0.586	0.058	0.154	0.125	0.097	0.093	0.015	0.021	0.030	0.014	0

Fig. 3. Relative Positions of MCCs in 2D Space Using MDS Visualization Algorithm (a) All MCCs (b) Without PH 301 and PH 302

A: PH 101, B: PH 102, C: PH 301, D: PH 302, E: Ceolus KG 801, F: Ceolus KG 802, G: Celex 101, H: Emcocel 50 M, I: Prosolv 50 M, J: Viva Pur 101, K: Pharmacel 101, L: Pharmacel 102, M: Comprecel M101 and N: Comprecel M102.

densities and extrusion-spheronisation parameters, $W_{710\,\mu m}$ and W_s (Table 1). The coarser of the 2 grades, M102, was found to be denser and relatively less sensitive to changes in water content. In addition, it required a smaller amount of water to make pellets of 710 μm size. These disparities in their physical properties did not result in a marked difference in their rheological profiles and their subsequent grouping.

For both M101 and M102, they were grouped together with the Pharmacel MCCs and Viva Pur. Within this group, no clear relationship or similarities between their physical characteristics could be observed. Their assigned grouping could be attributed to the combined effects from the interplay between their unique set of physical properties that result in their equivalent rheological profiles when moistened. In other words, the contributions from the individual property were balanced out, producing similarities in the rheological profiles obtained.

Particle size was not found to have an effect on the rheological profiles of the moistened MCC mass as indicated by M101 and its large particle size counterpart, M102 which were assigned into the same group, regardless of the stringency of clustering. This observation corroborated with the grouping results for Pharmacel 101 and Pharmacel 102.

Conclusion

The utility and value of MDS as a pre-formulation tool in the grouping of an expanded range of MCC grades was exemplified in this study. On its own, MDS provided a rapid visual comparison between the wide varieties of MCCs grades. However, its potential can be better harnessed when used in combination with other computational techniques. Our pro-

Table 3. Grouping of MCCs

Group	Level One	Level Two
	IT=0.9	IT=0.8
1	Avicel PH 102, Ceolus KG 801	Celex 101, Ceolus KG 801, Ceolus KG 802,[a] Avicel PH 102
2	Emcocel 50 M , Prosolv 50 M	Emcocel 50 M, Prosolv 50 M
3	Pharmacel 101, Pharmacel 102, Viva Pur, Comprecel M101,[a] Comprecel M102[a]	Pharmacel 101, Pharmacel 102, Viva Pur, Comprecel M101,[a] Comprecel M102[a]
—	Celex 101	Avicel PH 101
—	Ceolus KG 802[a]	Avicel PH 301
—	Avicel PH 101	Avicel PH 302
—	Avicel PH 301	—
—	Avicel PH 302	—

a) Additional MCCs included in this study.

posed pre-formulation tool which incorporates ANN, data clustering and MDS, was capable of assigning newly introduced or unknown MCCs into existing groups based on their torque rheological properties. The latter, was in turn, governed by the unique interactions between their physical properties. The sensitivity of this system was demonstrated by the different grouping result for Ceolus KG 802 when the levels of stringency for clustering were changed slightly even though their differences were believed to be minimal.

Other advantages of this pre-formulation tool include its simplicity of usage and flexibility in the type of data set used. By subjecting MCC grades to varying degrees of moisture, the rheological profiles of their wetted masses were able to reflect their differences and enabled their assignment into groups where the members possessed equivalent and comparable water–MCC interaction. This signifies a possible prediction of their performance in processes such as extrusion-spheronisation and confers flexibility or interchangeability in the selection of MCC grades without having to perform a laborious set of actual trial formulation studies. In addition to torque rheological profiles, other parameters or characteristics can also be used to group the MCC grades. This feature will undoubtedly be of great value to pharmaceutical sciences, especially in the field of pre-formulation.

Acknowledgements The authors wish to thank Asahi Kasei Chemical (Osaka, Japan) and Ming Tai Chemical Co. Ltd. (Taiwan, Republic of China) for their generous gifts of Ceolus KG 802, Comprecel M101 and M102.

References

1) Ebube N. K., Owusu-Ababio G., Adeyeye C. M., *Int. J. Pharm.*, **196**, 27—35 (2000).
2) Soh J. L. P., Chen F., Liew C. V., Shi D. M., Heng P. W. S., *Pharm. Res.*, **21**, 2361—2368 (2004).
3) Borg I., Groenen P., "Modern Multidimensional Scaling: Theory and Applications," Springer-Verlag, New York, 1997.
4) Barr G., Dong W., Gilmore C. J., *J. Appl. Crystallogr.*, **37**, 243—252 (2004).
5) Iyer R. M., Augsburger L. L., Pope D. G., *Pharm. Dev. Tech.*, **1**, 325—331 (1996).
6) Fielden K. E., Newton J. M., Rowe R. C., *Int. J. Pharm.*, **97**, 79—92 (1993).
7) Bains D., Boutell S. L., Newton J. M., *Int. J. Pharm.*, **69**, 233—237 (1991).
8) Heng P. W. S., Koo O. M. Y., *Pharm. Res.*, **18**, 480—487 (2001).
9) Robel A., "Proceedings of the 6th IEEE Workshop on Neural Networks for Signal Processing," Kyoto, Japan, 1996, pp. 190—199.
10) Meulman J., "A Distance Approach to Nonlinear Multivariate Analysis," DSWO Press, Leiden, 1986.
11) Cox T. F., Cox M. A. A., "Multidimensional Scaling," Chapman and Hall, London, 1994.