Available online at www.sciencedirect.com
SCIENCE DIRECT
Chemometrics and Intelligent Laboratory Systems 81 (2006) 21 – 28
Chemometrics and intelligent laboratory systems
www.elsevier.com/locate/chemolab

The use of principal component analysis and self-organizing map to monitor inhibition of calcium oxalate crystal growth by *Orthosiphon stamineus* extract

Saravanan Dharmaraj *, M. Amzad Hossain, Salman Zhari, Gam Lay Harn, Zhari Ismail

School of Pharmaceutical Sciences, University Science of Malaysia, 11800 Pulau Pinang, Malaysia

Received 8 December 2004; received in revised form 18 August 2005; accepted 5 September 2005
Available online 14 November 2005

Abstract

The modified Schneider's gel slide method is used to monitor the inhibitory effect of 50% methanol extract of *Orthosiphon stamineus* on the growth of calcium oxalate crystals. The images of crystals grown in the gel were captured at 1, 2, 3, 4 and 24 h with digital camera and image analysis was carried out. A total of nine variables relating to size and shape parameters were calculated and six were used for further analysis. Principal component analysis (PCA) and self-organizing map (SOM) were applied in the visualization of the size and shape distribution of the produced crystals using the following six variables: area, convex area, aspect ratio, equivalent diameter, roundness and full ratio. The results indicate that the modified gel slide method with the use of PCA and SOM to reduce the dimensionality of the data allowed an intuitive presentation of the differences in the studied inhibitors. The decrease in crystal growth for the extract and positive control of sodium citrate was clearly evident in the principal component scores and also on the location in the SOM. Evidence about the decrease in growth was also provided by an inhibition index of 0.290 for the extract and 0.255 for the positive control.

Keywords: Principal component analysis (PCA); Self-organizing map (SOM); *Orthosiphon stamineus*; Calcium oxalate

1. Introduction

There are many theories for the formation of kidney stones. One of them, the crystallization precipitation theory, implies that supersaturation of urine leads to precipitation of stone crystallites. These critical particles become entrapped and subsequent crystal growth follows. The process of aggregation and secondary nucleation leads to the formation of large crystals, which ultimately develop into kidney stones [1]. Another theory, the inhibitory theory, suggests that normal urine contains substances that inhibit the crystallization of calcium oxalate. Inhibitors of crystal growth block the growth of crystals and prevent stone formation [2].

Other than normal drugs, herbals are also commonly used to treat urinary stones and studies of such local Malaysian medicinal plants confirm the presence of compounds that show inhibitory effect in growth of calcium oxalate [3].

Among the methods to measure the inhibitory activity of crystals are photometry [4], turbidimetry [5], mixed suspension mixed product removal MSMPR [6], use of Coulter counter [7–9] and spectrophotometry [10]. Each of this method has some advantages but there is no ideal technique. A comparison of seven different methods to measure crystallization in urine were reported in a workshop in 1987 and the approach of crystallization in gels contained in wells of microtitre plate gave good efficiency and was suggested for basic research and clinical routine [11]. In a subsequent workshop [12], problems relating to crystallization measurements were discussed, for example, Coulter counter analysis was laborious, and optical techniques involving densitometry approaches required high crystal density. However, most of above techniques measured inhibition indirectly and not the effect on single crystals.

Nowadays, the use of image analysis offers a way to quantify the variations in crystal population. Shape and size can be characterized subsequent to the visualization of the crystals by light microscopy. The quantitative description of the morphology of the produced crystals is with the use of different

* Corresponding author. Tel./fax: +60 4 6563443.
E-mail address: saravanandharmaraj@hotmail.com (S. Dharmaraj).

0169-7439/$ - see front matter © 2005 Elsevier B.V. All rights reserved.
doi:10.1016/j.chemolab.2005.09.007

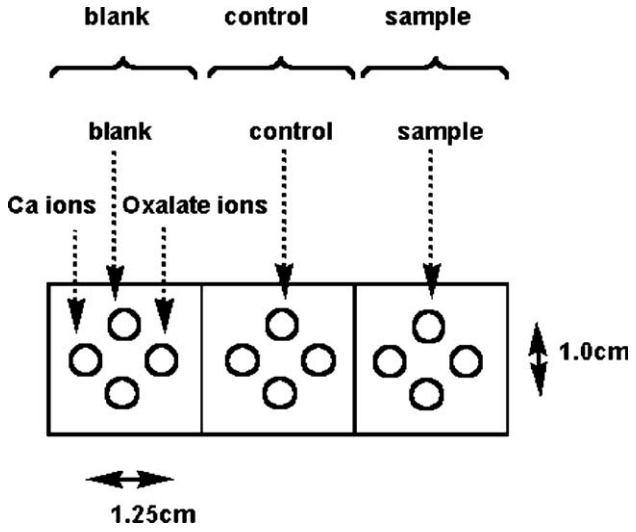

Fig. 1. A diagram of the gel slide.

shape and size descriptors. A simple method that gives a rapid insight into the variations of these descriptors is by using principal component analysis (PCA) to summarize all the shape and size parameters as carried out by Bernard-Michel et al. [13]. Another approach is the use of self-organizing map (SOM), which was reported by Laitinen and coworkers [14] as giving a better perception of the distribution compared to PCA.

In this study, a modified Schneider's gel slide method [4] was used to study the inhibition of calcium oxalate crystal growth by 50% methanol extract of leaf of *Orthosiphon stamineus* at concentration of 5000 ppm. A positive control of sodium citrate 10 ppm solution was also used to monitor the inhibition. A large set of data consisting of size and shape parameters was produced by image analysis while PCA and SOM were used to analyze the data. The objective of this work, other than to monitor the inhibitory effect of the plant extract on crystal growth, was also to study the feasibility of using PCA and SOM on image analysis data of modified gel slide method to monitor the effects on crystals.

2. Methods

2.1. Gel slide method

A solution of 4 ml 1% bacteriological agar was used to coat a microscopic slide that was partitioned into three equal areas. Four wells were punched into the agar on each side of the partition. Two wells were made 1.25 cm apart along the longer axis while two wells were made 1.0 cm apart along the perpendicular axis as shown in Fig. 1. Solutions of 10 μl 0.2 M calcium chloride and ammonium oxalate were pipetted into opposite wells along the longer axis. Solutions of 10 μl from 50% methanol extract or positive control consisting of sodium citrate were pipetted into the other two holes. The gel slides were placed in a moist chamber. The crystals formed were monitored at 1, 2, 3, 4 and 24 h under microscope and images were captured with a digital camera mounted onto the microscope. The slides were analyzed in triplicate and six images were captured for each subsection of blank, positive control and sample.

The optical microscope (Leica MZ6, Leica Microskopie und Systeme, Germany) was connected to image analysis (IA) software (Leica Qwin, Leica Imaging Systems, Cambridge, England), which was used to calculate the size and shape parameters of the formed calcium oxalate crystals. Nine parameters of shape and size were measured of which only six were used for PCA. The image analysis system measures many very similar kinds of parameters, the ones with high correlation values were omitted to eliminate factors that do not bring any additional information to the analysis of the results. Parameters of length, breadth and perimeter were not used.

The six chosen parameters for further analysis were: area, convex area, aspect ratio, equivalent diameter, roundness, and fullness ratio, and they are presented in Table 1 [14]. The size parameters are area, convex area and equivalent diameter while the other three represent shape parameters.

2.2. Multivariate analysis

Two methods of multivariate analysis were used to obtain a description of the various data sets. PCA and SOM were used to enable the lowering of the dimensionality in the multivariate data.

2.2.1. Principal component analysis

PCA has been described as the basic workhorse of multivariate analysis and among the most popular linear projection method [15]. The measured image analysis data consisting of mean values of the six parameters for the obtained crystals were evaluated using PCA employing SPSS program. PCA is a data visualization method that is useful for observing groupings within multivariate data. Data is represented in n dimensional space, where n is the number of variables, and is reduced into a few principal components, which are descriptive dimensions that describe the maximum variation within the data. The principal components can be displayed in a graphical fashion as a 'score' plot. This plot is useful for observing any groupings or trend in the data set [16].

2.2.2. Self-organizing map

SOM was used to train and visualize the crystal data, which consisted of values of six parameters measured from the crystals. The SOM map is available as a Matlab toolbox in a

Table 1
Description of parameters used in the analysis

Parameters	Description [14]
Area	The apparent area of the crystals
Convex area	The area of the polygon circumscribing the feature formed by tangents to its boundary
Aspect ratio	The ratio of particle length divided by its breadth
Equivalent diameter	Equivalent circle diameter—i.e. the diameter of a circle having the same area as the feature
Roundness	A shape factor which gives a minimum value of unity for a circle (round particles get a value of 1, and other particles get values larger than 1)
Fullness ratio	A shape factor equal to the square root of the ratio of area to circumscribed area

public domain in the Internet [17] and the map was trained using a Pentium 3 computer. SOM is basically an unsupervised neural network and has two layers, the input layer and output layer. The input layer is one-dimensional, whereas the output layer is usually two-dimensional and often arranged in a way that each unit or neuron is the neighbor of another six; which is why they are represented as hexagons. A neighborhood arrangement is maintained in the SOM algorithm where nearby map neurons have similar profiles after training. Every neuron has a parametric reference vector or prototype vector and the whole input data set in the form of input vectors is presented for training according to the batch version of Kohonen's SOM training algorithm and each observation is projected onto a winning node [18,19]. The neuron with the prototype vector having the shortest distance from the presented vector is the winning node. The vectors of the winner and its neighboring nodes are modified following the training to represent the input signals in a topology preserving fashion. Finally, the SOM net can project data from an n-dimensional space to a usually two-dimensional grid of neurons and thus facilitate a better visualization of the data.

2.3. Inhibitory effect

Statistical test of ANOVA by SPSS was carried to determine if there were statistically significant differences in the area of the crystals between blank, control and extract. The effects of the extract and positive control on the in vitro growth of calcium oxalate crystals were also compared with the blank by using the inhibition index (I). Absence of inhibition is indicated by I equals 0, whereas complete inhibition is shown by I equals 1 [20].

Inhibition index = 1 − (As/Ac), where As = area of calcium oxalate crystals in the presence of tested sample and Ac = area of calcium oxalate for the corresponding blank.

3. Results and discussion

3.1. Image analysis measurements

The monitoring of crystal growth was conducted with modification to Schneider's method [4]. In the Schneider's method, 3 ml of agar was used to coat the microscopic slide and technique of photometry, which produced a single parameter used to monitor the inhibition. In our work, we used 4 ml of agar and monitored inhibition using image analysis. However, the major modification was the distance between the two longitudinal wells that was reduced from 2.5 cm to 1.25 cm in our work. This was to enable comparison of blank, control and sample on the same slide, which was not possible in the original method.

The crystal size distribution obtained for the blank, control of sodium citrate and sample of *O. stamineus* extract at 1, 2, 3, 4 and 24 h are shown in Fig. 2.

Overall, the size distribution shows that the smallest crystals are most numerous for all three; blank, control and sample, at all time periods studied. There is an increase in the size of the biggest crystals with time, especially more apparent for blank. This increase in size is less obvious for control and sample between 1 and 24 h. When the blank, control and sample data are compared at 24 h, it is apparent that mean size distribution is higher for blank compared to control and sample. This is shown by the normal curve that peaks at a higher area value for the curve for blank when compared to control and sample. The normal curve also declines more gradually for the blank when compared to control and sample. The growth of calcium oxalate crystals with time can also be observed by plotting the mean area with time as shown by the histogram in Fig. 3. The decline in mean area for positive control and extract is due to increase in number of smaller crystals for them at 24 h compared to the preceding period. This could be due to delayed nucleation of calcium oxalate crystals.

3.2. PCA

The initial step was to check the appropriateness of using factor analysis method of PCA to summarize all the morphological variation produced by the image analysis data for the calcium oxalate crystals. The measure of sampling adequacy (MSA) is a measure to quantify the degree of intercorrelations among the variables and is used to justify the use of PCA. This index ranges from a value of 0 to 1, reaching 1 when each variable is predicted without error by the other variables. The PCA analysis for the crystal data gave a MSA value of 0.655 that justified the use of this factor analysis. The Bartlett test of sphericity, a statistical test for the presence of correlations among the variables, is another such measure and it also justified this analysis [21].

PCA was found capable of giving a rapid insight into the variations of the size and shape variables. The principal components are determined on the basis of maximum variance criterion. Each subsequent principal component describes a maximum of variance that is not modeled by the former component. According to this, most of the variance is contained in the first principal component and then in the second principal component there is more variance than the third. Generally, the first two components are used and they are capable of explaining a high percentage of variance in the data.

There are various criteria for the number of factors to extract and in this analysis the latent root criteria is used. Therefore, only factors having latent roots or eigenvalues greater than 1 are considered significant and this gave two factors. Another usual criterion is to obtain the percentage of explained variance that is greater than 90% and the two factors also fulfilled this condition [22]. In most cases, rotation of the factor improves the interpretation by reducing some of the ambiguities that often accompany initial unrotated factor solution and with this data, rotation method of varimax with Kaiser normalization gave a better perspective.

Correlations and importance of the size and shape variables can be observed from the rotated principal component loadings as shown in Fig. 4. The size of the loadings for a particular variable in the considered principal component is a measure of the importance of that variable or feature for the principal

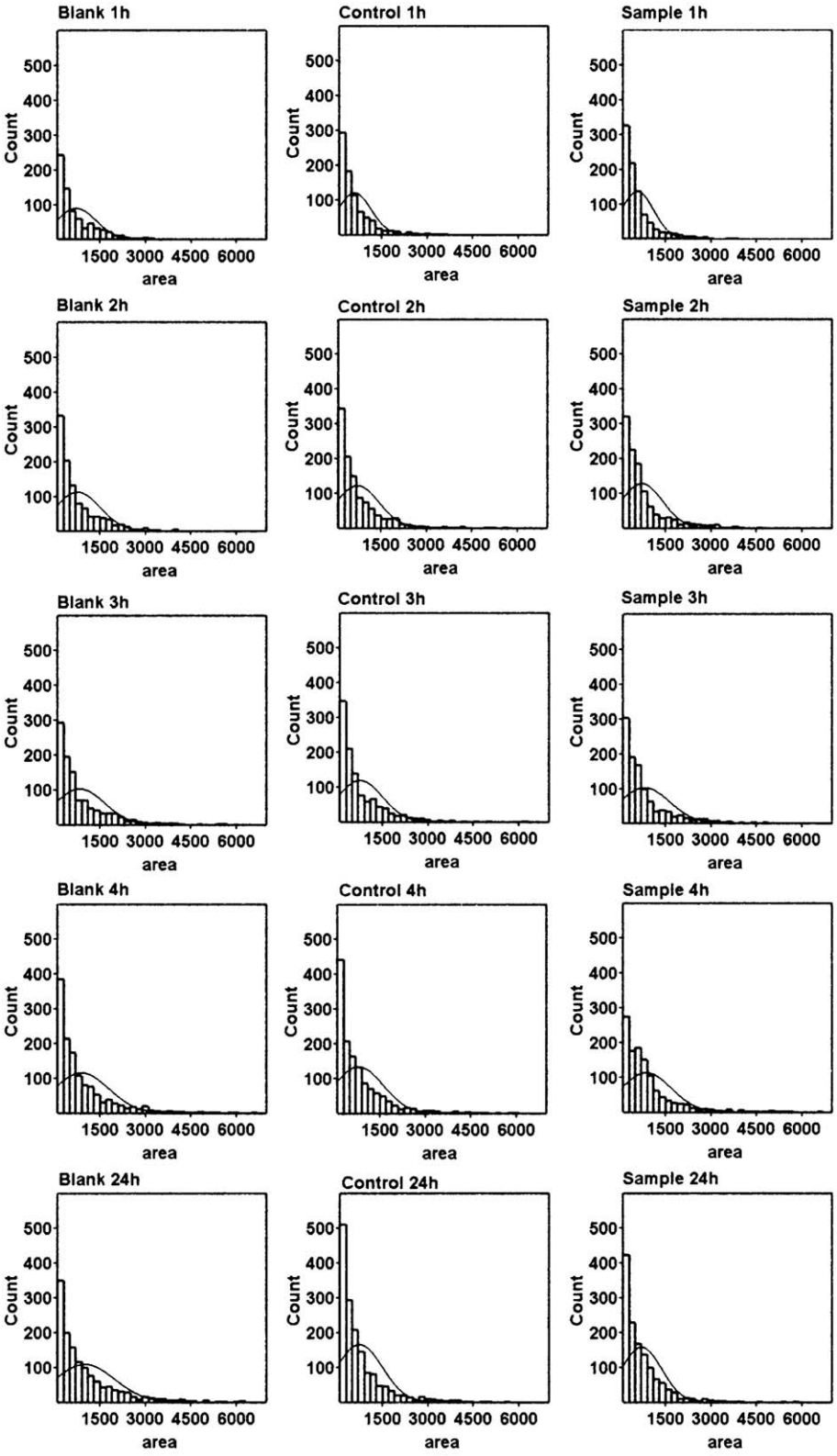

Fig. 2. The histogram of the crystal size distribution.

component concerned. It can be noticed that the size parameters of equivalent diameter, area and convex area are highly correlated among them in the rotated component plot and load highly on the first principal component with values of 0.959, 0.949 and 0.925 respectively. The shape parameter of full ratio loads highly (−0.930) on the second principal component and has negative correlation to roundness and aspect ratio. Shape parameters of roundness and aspect ratio are highly correlated among them and also load highly on the second component with values of 0.922 and 0.925 respectively.

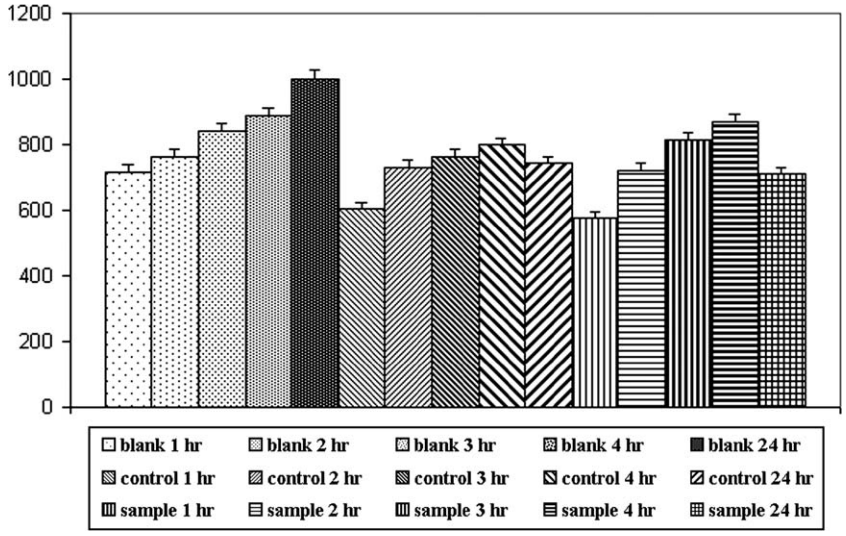

Fig. 3. The growth profile for blank, control and sample at 1, 2, 3, 4 and 24 h. The error bars indicate the values for S.E.M. Unit in μm^2.

Basically, first principal component is related to size parameters and second principal component to shape parameters.

The percentage of explained variance gives an idea of how much variance can be explained by the use of the principal component model, which shows how good the model is. The fraction of total explained variance Se^2 is calculated from the ratio of the sum of d-important eigenvalues (λ) and the sum of all p eigenvalues by the use of:

$$S_e^2 = \frac{\sum_{i=1}^{d} \lambda}{\sum_{i=1}^{p} \lambda}$$

The eigenvalues for the rotated six components in this analysis are 3.001, 2.858, 0.124, 0.012, 0.004, and 0.000. PCA indicated that the rotated first component explained about 50.0% of the variance in the crystal data. The rotated second component explained about 47.6% of the variance. The two components combined explained 97.7% variance in the crystal data. The two rotated principal components form a plane in the original image analysis data space and are shown in Fig. 5. The crystal data for blank at 1, 2, 3, 4 and 24 h shows that there is an increase in value of first principal component score with time for blank crystals, which means an increase in size as principal component one corresponds to size parameters. The crystal data also show an increase in this score for control and sample data between 1 and 4 h. However, the first principal component score at 24 h is much reduced for control and sample when compared to their respective 4 h data. There is a great decrease in this score for control and sample when compared to blank at 24 h. Basically, this means a decrease in size due to the inhibitory effect on crystal growth by positive control of sodium citrate and 50% methanolic extract.

There seems to be no corresponding pattern in decrease or increase for principal component two when comparing for blank, control and sample at the time periods studied. However, there is an increase in value for the second component for blank and sample at the 3 h period.

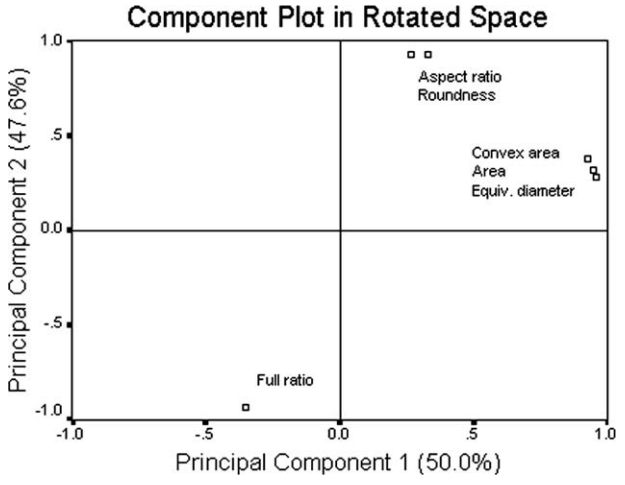

Fig. 4. Plot of principal component loadings. Area, equivalent diameter and convex area are size parameters for crystals and load highly on first principal component, whereas aspect ratio, roundness and fullness ratio are shape parameters. The magnitude of shape parameters for the second component show that they load highly on it.

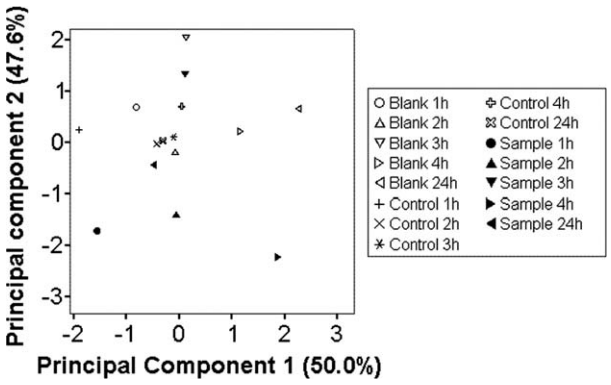

Fig. 5. Principal component scores for the various test samples. Principal component 1 is related to size parameters and principal component 2 to shape parameters. An increase in principal component 1 score corresponds to increase in size of crystal.

3.3. SOM

The training data was used to form a matrix by combining the image analysis data of the crystals from the measurements. The dimensions of the matrix were $17{,}017 \times 6$. The matrix included the data points of the six parameters that described the 17,017 crystals measured from the set of blank, control and sample at 1, 2, 3, 4 and 24 h. The size of the SOM that was used for the model particles was 30×22. The training of the net for the model particles took 22 s.

A common visualization technique for trained SOM is distance matrices; a matrix of distances between neighboring map nodes. The unified distance matrix or U-matrix represented the distances of the six parameters in a single map, whereas separate maps were also used for each parameter. Both the U-matrix and maps for the six parameters are shown in Fig. 6. The color of the node indicates the level of the individual variable on the specific region of the map. The high values are indicated with red color and low values with violet color.

It can be noticed that crystals with largest values for size parameters (area, equivalent diameter and convex area) are located at the upper region with the maximum values at the upper right. The crystals with the large values for shape factor roundness and aspect ratio are located in the upper left with the maximum values for aspect ratio being located below slightly compared to the ones for roundness parameter. The crystals with the lowest values for fullness ratio are located at the top right side of the map.

Fundamental to the SOM technique is the location of a given data on the map or the response of the map to an input data. The location is known as best-matching unit (BMU) and for a particular crystal data is the map unit with prototype vector closest to the input vector. The organization of the crystal data on the map for blank, positive control and sample at 1, 2, 3, 4 and 24 h is shown by locations of the corresponding BMUs in Fig. 7. It is observed that from BMU for blank crystals (represented by white hexagons), the crystal data is clustered at the lower part of the map at 1 h and with increasing time the data is clustered towards the top parts. Size of hexagons indicates number of crystals with similar magnitude for the parameters studied. There are numerous white hexagons of big size at the bottom at 1 h blank but these are less in the 24 h blank data. The big white hexagons are situated in the upper right corresponding to bigger size parameters for the 24 h blank crystals. Overall, there seems to be greater number of small crystals at earlier period and a corresponding increase in crystal size with time. This increase in size is less obvious for control (yellow) and sample (magenta) between 1 and 24 h. When the blank, control and sample data is compared at the period of 24 h, it is apparent that there are less white hexagons at the bottom compared to yellow and magenta hexagons. However, there are more white hexagons at the upper right in comparison to yellow and magenta hexagons. Taken together, this means that there is a corresponding decrease in crystal size for control and sample when compared to blank at 24 h.

Regarding shape parameters there is not much difference in the blank, control and sample at the time period studied as found for size parameters. This is apparent as the crystals in the image were generally of similar type that is mostly calcium oxalate dihydrate.

The SOM summarizes effectively large sets of particle size and shape parameters. The advantage of SOM as reported is the ability to efficiently visualize the distribution of crystals on the map.

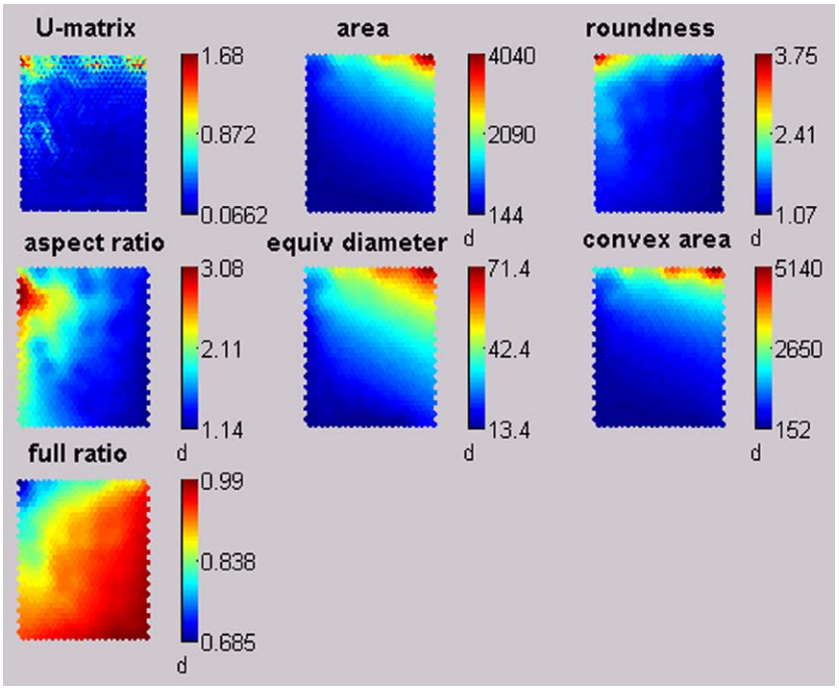

Fig. 6. The U matrix and the variable information for the particles.

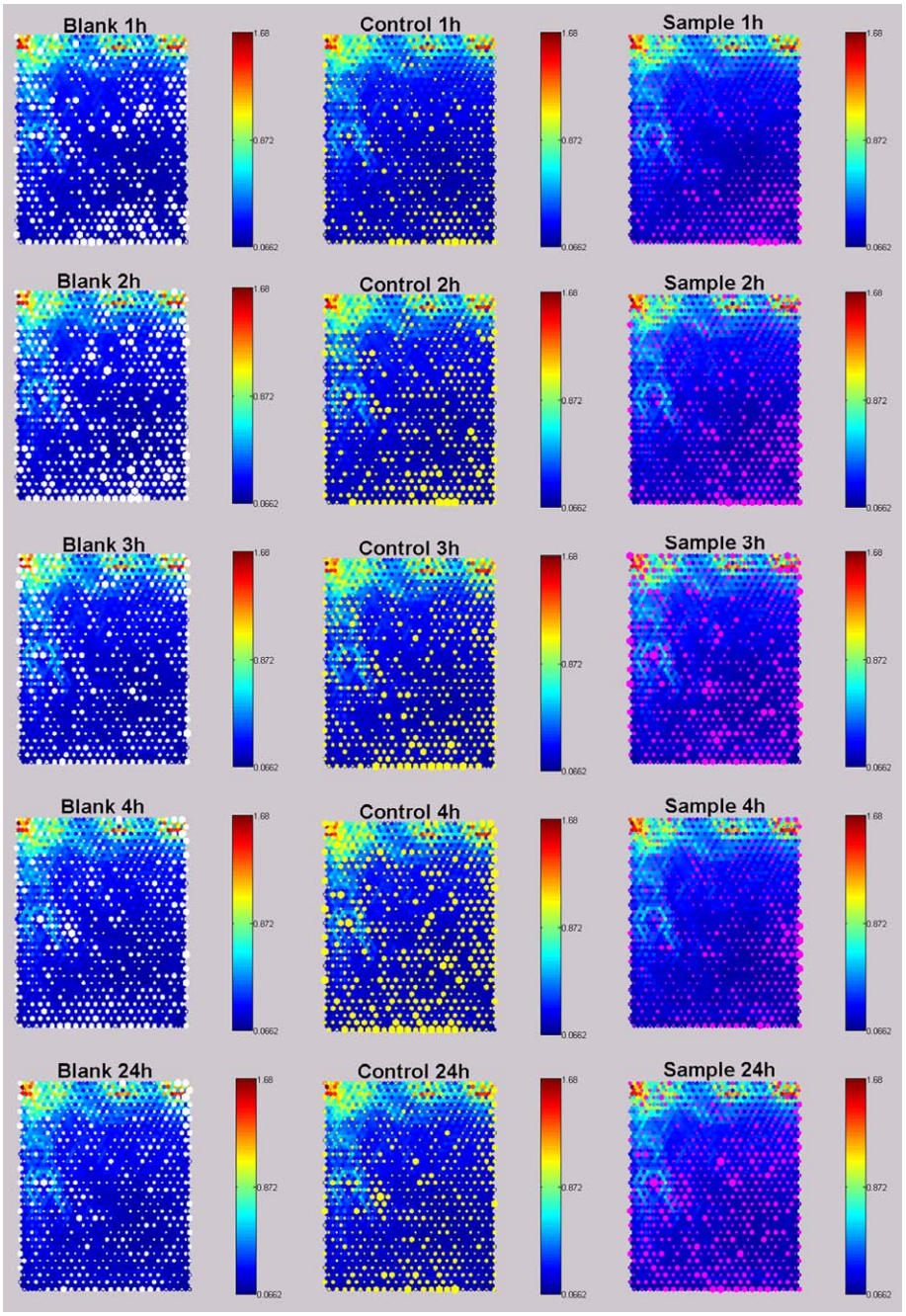

Fig. 7. The crystal data on the map for blank (white), positive control of sodium citrate 10 ppm (yellow) and sample of 5000 ppm 50% methanolic extract of *Orthosiphon stamineus* (magenta) at 1, 2, 3, 4 and 24 h. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

3.4. Inhibition of crystal growth

Both PCA and SOM complement each other and further suggest that there is inhibition due to both extract and positive control. The principal component loadings in PCA and location in SOM suggest that size variables have high correlation between them. This allows us to use the whole area data for individual crystals for ANOVA by SPSS to determine if the differences were statistically significant. As the number of crystals measured in each treatment varied, SPSS calculated and used a harmonic mean sample (crystals) size of 1092. Applying the test of Tukey HSD on the blank data showed that there was a statistically significant difference in the crystal size for blank except when compared to the adjacent time period. This can be seen by the groupings of increasing size: a) blank 1 h and blank 2 h, b) blank 2 h and 3 h, c) blank 3 h and blank 4 h and d) blank 4 h and 24 h. More importantly, there were differences between the positive control and extract at 24 h when compared to the corresponding blank, indicating inhibition of growth. The results of the Tukey test by SPSS were confirmed by using a slight modification of the calculation as described by Fowler et al. [23]. It showed slight differences with the results by SPSS but confirmed the growth of blank crystals with time and also the inhibitory effect of

extract and positive control on crystal growth at 24 h. The effect on crystal growth by inhibitors was also suggested in Schneider's paper [4] but the photometric analysis was not able to verify this and furthermore quantification on size of crystals in images was not done in their work.

The inhibition indexes were calculated using the 24 h data and the value for the extract was 0.290, whereas for sodium citrate 10 ppm was 0.255. Sodium citrate is a known inhibitor of calcium oxalate crystals and its inhibitory effect was also noted in an earlier study [4].

4. Conclusions

The modified gel Schneider's method with image analysis method combined with multivariate techniques of PCA and SOM was capable of monitoring the calcium oxalate crystal growth. The use of PCA enabled easy interpretation of the large data sets produced by reducing the dimension of the data. The score for principal component one of the rotated data matrixes is associated with size parameters, whereas the score for principal component two is associated with shape parameters. The principal component one score of the blank, control and samples at 24 h enabled us to monitor the crystal inhibition properties of an extract of *O. stamineus*. A positive control also gave similar inhibition. SOM was able to complement PCA by providing a visual perception of the whole data set as location of the data on the map can be related to size and shape parameters. Calculation of the inhibition index by using area data confirmed the inhibitory effect of both the 5000 ppm 50% methanolic extract of *O. stamineus* and 10 ppm sodium citrate. Overall, the gel slide method in combination with image analysis with PCA and SOM can be used as a screening method for compounds or herbals on calcium oxalate crystal growth inhibition. Principal component scores in PCA or location in the SOM shows effects of the tested samples on growth or on morphology of crystals.

Acknowledgement

An Intensifying Research Priority Areas (IRPA) grant from the Ministry of Science, Technology and Environment (MOSTE), Malaysia supported the study. Saravanan Dharmaraj wishes to thank MOSTE for a PhD scholarship.

References

[1] M.W. McDonald, M.L. Stoller, Geriatrics 52 (1997) 39.
[2] G. Bihl, A. Meyers, Lancet 358 (2001) 651–656.
[3] I. Zhari, R.M. Sahabudin, N. Ismail, M.N. Ahmad, H.M. Sirajinisa, S.M.A. Raihan, A. Temumen, Mal. J. Sc. 18 (1999) 13–15.
[4] H.J. Schneider, W. Bothor, W. Berg, R.H. Borner, M. Jakob, Urol. Int. 38 (1983) 33–38.
[5] Th. Briellmann, F. Herring, H. Seiler, G. Rutishauser, Urol. Res. 13 (1985) 291–295.
[6] T. Bretherton, A. Rodgers, J. Cryst. Growth 192 (1998) 448–455.
[7] P.K. Grover, R.L. Moritz, R.J. Simpson, R.L. Ryall, Eur. J. Biochem. 253 (1998) 637–644.
[8] O. Miyake, K. Yoshimura, T. Yoshioka, T. Koide, A. Okuyama, Urol. Res. 26 (1998) 209–213.
[9] K. Suzuki, K. Kawamura, R. Tsugawa, Scanning Microsc. 13 (1999) 183–189.
[10] B. Hess, C. Jost, L. Zipperle, R. Takkinen, P. Jaeger, Nephrol. Dial. Transplant. 13 (1998) 2241–2247.
[11] Anonymous, Urol. Res. 16 (1988) 137–142.
[12] Anonymous, Urol. Res. 17 (1989) 143–151.
[13] B. Bernard-Michel, M.N. Pons, H. Vivier, S. Rohani, Chem. Eng. J. 75 (1999) 92–103.
[14] N. Laitinen, J. Rantanen, S. Laine, O. Antikainen, E. Rasanen, S. Airaksinen, J. Yliruusi, Chemometr. Intell. Lab. Syst. 62 (2002) 47–60.
[15] K.H. Esbensen, Multivariate Data Analysis—In Practice, 5th ed., CAMO Process AS, Oslo, 2002, pp. 19–104.
[16] D.L. Massart, B.G.M. Vandeginste, S.N. Deming, Y. Michotte, L. Kaufman, in: B.G.M. Vandeginste, L. Kaufman (Eds.), Chemometrics: A Textbook, Elsevier, Amsterdam, 1988, pp. 339–370.
[17] E. Alhoniemi, J. Himberg, J. Parhankangas, and J. Vesanto, SOM Toolbox 2 for Matlab, http://www.cis.hut.fi/projects/somtoolbox/, Helsinki University of Technology, Finland, 2000.
[18] T. Kohonen, in: Self-Organizing Maps, Series in Information Sciences, vol. 30, Springer, Heidelberg, Germany, 1997, p. 426.
[19] J. Vesanto, Using SOM in data mining. Licentiate's thesis. Helsinki University of Technology, 2000.
[20] M.K. Li, D.K.Y. Shum, S. Choi, Urol. Res. 15 (1987) 75–77.
[21] J.F. Hair, R.E Anderson, R.L. Tatham, W.C. Black, Multivariate Data Analysis, 5th ed., Prentice Hall, New Jersey, 1998, pp. 87–138.
[22] M. Otto, Chemometrics: Statistics and Computer Application in Analytical Chemistry, Wiley-VCH, Weinheim, Germany, 1999, pp. 119–173.
[23] J. Fowler, L. Cohen, P. Jarvis, Practical Statistics for Field Biology, 2nd ed., John-Wiley & Sons, Chichester, England, 1998, pp. 179–209.