Pharmaceutical Research, Vol. 23, No. 9, September 2006 (© 2006)
DOI: 10.1007/s11095-006-9037-y

Research Paper

Near-Infrared Imaging for Studying Homogeneity of Protein-Sugar Mixtures

Nataša Jovanović,[1] Ad Gerich,[2] Andréanne Bouchard,[3] and Wim Jiskoot[1,4]

Received February 2, 2006; accepted April 13, 2006; published online August 12, 2006

Purpose. To investigate the applicability of near-infrared (NIR) imaging for assessing the homogeneity of dried protein-sugar formulations.
Methods. Physical mixtures of lysozyme and trehalose in different ratios were prepared and analyzed by near-infrared (NIR) imaging with a spatial resolution of 10 or 40 μm. To define and select the best imaging strategy, besides visual inspection of the images, several approaches for data processing were tested: single wavelength intensity, peak/height ratio of two specific wavelengths, correlation coefficient with a reference spectrum and principal component analysis (PCA). In order to relate the contrast directly to concentration differences of lysozyme and trehalose, quantitative models were created based on correlation coefficient and partial least squares (PLS) regression. The selected imaging method was applied to compare the homogeneity of a supercritical fluid (SCF) dried and a freeze-dried lysozyme-trehalose mixture.
Results. All tested methods confirmed each other and showed spatial heterogeneity in the lysozyme and trehalose contents of the physical mixtures. However, multivariate data processing methods (correlation coefficient and PCA/PLS) resulted in more distinct contrasts than univariate approaches (single wavelength analysis) and allowed a quantitative estimation of the homogeneity. As shown by NIR imaging in combination with the correlation coefficient or the PLS method, the SCF dried lysozyme-trehalose formulation was at least as homogeneous as its lyophilized counterpart, at 10 μm pixel size resolution.
Conclusions. NIR imaging is a useful tool for studying the homogeneity of dried protein-sugar formulations.

KEY WORDS: homogeneity; near-infrared imaging; phase separation; protein; supercritical fluid drying; trehalose.

INTRODUCTION

Producing stable protein formulations has been the focus of many investigations during the past few decades (1,2). A common approach to stabilize proteins is to dry them, e.g., by freeze drying, spray drying or supercritical fluid (SCF) drying (1,3,4). However, during drying and on storage of the dried products protein degradation can occur (1,5,6). Therefore, the addition of sugars or other stabilizing excipients is generally required (1,7,8).

Regarding the quality of the dried product, a homogeneous distribution of protein and sugar can be expected to improve stability. Exclusion of protein from sugars could lead to aggregation during storage (1,9,10) and phase separation has been reported to be responsible for failure in protein stabilization (10–13). Phase separation may be difficult to detect with conventional methods such as differential scanning calorimetry (DSC) because of relatively low thermal signals, or scanning electron microscopy (SEM) which relies on visual observation. Near-infrared (NIR) imaging could be the method of choice (14) because it can determine spatial distributions of different components in pharmaceutical samples based on their specific NIR bands. NIR imaging is a fast, non-invasive technique, which enables monitoring of large sample areas and requires relatively small amounts of sample. The added value of NIR imaging is that besides visual inspection of images it is possible to locally quantify the concentrations and thus measure (in)homogeneity.

Recent improvements in the technical characteristics of NIR imaging equipment have resulted in the possibility to use NIR imaging for fast scanning of powder blend homogeneity (14,15). By recording an image of a sample, a picture based on NIR absorption is created. Every pixel contains a

[1] Department of Pharmaceutics, Utrecht Institute for Pharmaceutical Sciences (UIPS), Utrecht, The Netherlands.
[2] Department of Pharmaceutics, N.V. Organon, Oss, The Netherlands.
[3] Process and Energy Laboratory, Delft University of Technology, Delft, The Netherlands.
[4] Department of Drug Delivery Technology, Leiden/Amsterdam Center for Drug Research (LACDR), Leiden University, Leiden, The Netherlands.
To whom correspondence should be addressed. (e-mail: a.gerich@organon.com)

ABBREVIATIONS: DSC, differential scanning calorimetry; NIR, near infrared; PAT, process analytical technology; PCA, principal component analysis; PLS, partial least squares regression; SCF, supercritical fluid; SD, standard deviation; SEM, scanning electron microscopy; SG, Savitsky Golay.

NIR spectrum. By focusing on the differences in NIR spectra, every pixel can be classified as a certain component or mixture of more components. In principle NIR imaging will provide information about the surface of the sample, but due to a certain penetration depth (dependent on the wavelength) a top layer will be analyzed. A typical pixel size for NIR imaging is 40 or 10 μm. The penetration depth, defined as DP_{50}, varies between 180 μm at 1,200 nm, 50 μm at 1,600 nm and 30 μm at 2,400 nm (16). At the given depths, 50% of the information (signal) is still retrieved. Based on the variation in penetration depth and pixel size, one pixel in a NIR image will provide information of a volume of about 10 pl for a 10 μm pixel size and about 160 pl for a 40 μm pixel size (for an average penetration depth of 100 μm).

In this work we investigated the possibility to use NIR imaging for studying the homogeneity of dried protein-sugar mixtures. Firstly, physical lysozyme-trehalose mixtures were prepared and measured by NIR imaging. Several data processing approaches were compared to establish the optimum imaging strategy. With the selected strategy we studied the homogeneity of SCF dried and freeze-dried lysozyme-trehalose formulations. Our data show that NIR imaging can be a useful tool for measuring the degree of homogeneity of dried protein-sugar formulations.

MATERIALS AND METHODS

Materials

Hen egg white lysozyme (~70,000 U/mg) was purchased from Fluka (Buchs, Switzerland) and trehalose (crystalline, dihydrate from *Saccharomyces cerevisiae*) was acquired from Sigma-Aldrich (Steinheim, Germany). All other chemicals used were obtained from commercial suppliers and were analytical grade.

Preparation of Protein-Sugar Mixtures

Different proportions of pure lysozyme and trehalose, previously ground to obtain powder (with particle size between 5–35 μm), were mixed in a mortar, without further grinding in order to create relatively inhomogeneous samples. The lysozyme content in the mixtures was 0, 10, 50, 80 and 100% (w/w). These mixtures were used to determine a suitable imaging strategy and to make a calibration curve for estimating the average concentration of lysozyme and trehalose in unknown samples.

Test samples were prepared by drying a solution containing 5% lysozyme and 5% trehalose (protein-to-sugar ratio 1:1 (w/w)). The solution was dried by a SCF drying process described before (7), or by freeze-drying, in a ZIRBUS sublimator 400, according the procedure reported previously by Zografi and Saleki-Gerhardt (17). The corresponding water contents were 3.1 and 0.8% (w/w) for the SCF dried and the freeze-dried sample, respectively.

Sample Pre-Treatment

In principle no sample preparation is necessary for NIR imaging. However, using powder samples as such can result in a relatively large variation between spectra caused by scattering and specular reflectance of individual particles (unpublished observations). The larger the particles are, the more apparent this effect will be. Additional grinding of powders in a mortar would reduce these unwanted effects and improve the quality of the image, but would also affect the sample itself, particularly the homogeneity of the sample. Therefore, it was decided to press a tablet of each sample in order to create a more defined and suitable surface area. Pressing a tablet will also affect the sample, but it is assumed that the homogeneity of the sample as a tablet will still be representative for that of the powder.

Data Acquisition and Data Pre-Treatment

All images were recorded with a Sapphire NIR Imager (Spectral Dimensions) using Sapphire Go Software (version 1.0). Magnifications corresponding to a 10 × 10 μm pixel or 40 × 40 μm pixel size were used. Spectra were recorded between 1,200 and 2,400 nm with 4 nm increments using 16 replicates. Before data handling, the original images were truncated to 160 × 160 pixel images (total 25,600 pixels). All raw reflectance spectra were converted using Isys 3.1 software (Spectral Dimensions) into absorption spectra and normalized: mean centered and scaled to unit variance by spectrum. After normalization the second derivative of each spectrum using Savitsky Golay (SG) filters (9/3) was calculated.

Imaging Strategy

The reference samples (physical mixtures of lysozyme and trehalose) were used to determine a suitable imaging strategy. Within an image a high contrast between the two components is necessary to investigate homogeneity. The contrast of an image can be based on a large number of spectral properties. The following strategies were chosen to be investigated in more detail: (1) the intensity of a single wavelength, specific for trehalose or lysozyme; (2) the peak-height ratio of two wavelengths, specific wavelengths for trehalose and lysozyme; (3) the correlation coefficient with a reference spectrum; (4) the score value of a principal component from a principal component analysis (PCA). If PCA describes the lysozyme–trehalose variation well, partial least squares regression (PLS) can be used to create a quantitative regression model. The first two approaches are based on specific wavelengths, whereas the last two are based on the entire spectra. If a clear and strong absorption band occurs for a specific component it is normally possible to make use of a single wavelength based contrast image to show the spatial distribution of that specific component. If the intensity at this wavelength is also dependent on other (physical or chemical) properties, the use of a reference wavelength is recommended as used in peak-height ratio based contrast images. The correlation coefficient based contrast image is based on the correlation coefficient of the spectrum of a single pixel with a reference spectrum (i.e., the average spectrum of an entire image, 160 × 160 pixels), like pure lysozyme or trehalose. The correlation coefficient is very sensitive to spectral variations anywhere in the spectral wavelength range. In principle, the correlation coefficient is not used for quantitative regression, but a known relation between concentration and correlation coefficient can pro-

vide a quick and simple estimation of the concentration of any component of which a reference spectrum is available. Both lysozyme and trehalose can be used as a reference spectrum, which results in independent estimations of lysozyme and trehalose. The correlation coefficient is calculated according to Eq. (1):

$$correlation\ coefficient = \frac{\sum_i x_i \cdot y_i}{\sqrt{\sum x_i^2 \cdot \sum y_i^2}} \quad (1)$$

where x_i = response at wavelength i of spectrum x (reference) and y_i = response at wavelength i of spectrum y (sample).

PCA is a very common data analysis tool for large amounts of highly co-linear data (18,19). Similar to the correlation coefficient, PCA is based on the entire wavelength range. PCA is a multivariate projection method designed to extract and display the systematic variation in a data matrix. PCA is used to represent a multivariate data set as a low dimensional plane, usually consisting of 2 to 5 dimensions (principal components) such that an overview of the data is obtained. Principal components are new uncorrelated linear functions of the original variables. The relation between the variation described by a principal component and the original variables is given by loadings or loading weights for each original variable. If for a certain variation, described by a specific principal component, loading values of the original variables are high, then that particular original variable is important for that variation. Less important or redundant variables show low loading values. The first principal component describes the largest systematic variation in the data set, the systematic variation described by the following principal components decreases with every component. The score of each sample (spectrum) for a specific principal component is related to the relative position of that sample in the total data set, for the particular variance described by the accompanying principal component.

Contrast images can be created based on the score value of each spectrum within an image of a specific principal component, which describes a specific spectral variation. In contrast with PCA models which describe variation qualitatively, PLS regression is a common tool for multivariate quantitative models (18). PLS is a regression extension of PCA, which is used to relate two data matrices, X (spectra) and Y (reference values), with each other. In analytical chemistry PLS is mainly used for multivariate calibration. Since the whole spectra are used in the modeling, the multivariate PLS model results in better predictive precision, and also much improved selectivity in comparison with traditional univariate calibration.

RESULTS AND DISCUSSION

Imaging Strategy

To define an appropriate imaging strategy, hand made physical mixtures were used to test several data processing methods. Selected methods were further investigated and compared in order to define the best way to quantitatively estimate the degree of homogeneity. Also, a suitable spatial resolution was selected.

Selection of Data Processing Method

The average second-derivative NIR spectrum of lysozyme and trehalose was calculated in order to investigate wavelength areas corresponding to specific absorption bands of lysozyme and trehalose (Fig. 1). A specific absorption band for trehalose at 1,432 nm is clearly present. This absorption band corresponds to the first overtone of the O–H stretch (20,21). Some other, less intense, specific wavelengths can be assigned to trehalose as well, but they show overlap with the lysozyme spectrum. The lysozyme spectrum also contains specific bands: at 1,496 nm (N–H

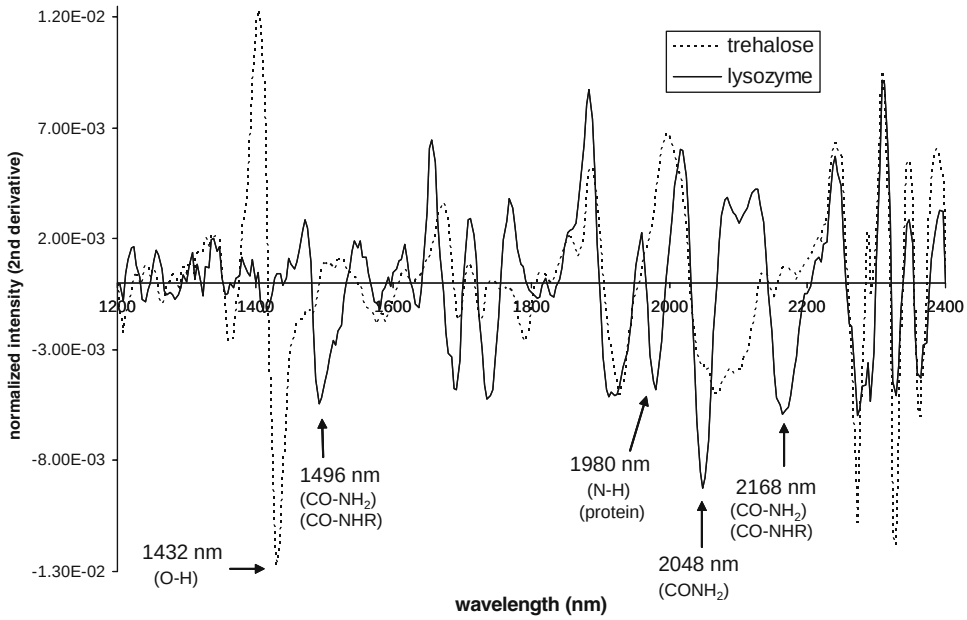


Fig. 1. Average second derivative (SG 9/3) NIR spectra of lysozyme and trehalose, after normalization.

stretch first overtone, CO–NHR), 1,980 nm (N–H anti-symmetric stretch + amide II combination band), 2,048 nm (N–H symmetric stretch + amide II combination band) and 2,168 nm (2 × amide I + amide III combination band). Regarding the shape of the trehalose spectrum, the absorption bands of lysozyme at 1,496 and 2,168 nm will probably be most suitable to distinguish between lysozyme and trehalose.

Single Wavelength Based Images

In Fig. 2 images of the lysozyme-trehalose reference mixtures (hand made physical mixtures) are given based on the specific wavelengths of trehalose (1,432 nm) and lysozyme (1,496 and 2,168 nm). In the first row of Fig. 2 the contrast of the images is based on the intensity at 1,432 nm, a higher intensity corresponding to a darker contrast. As expected, 100% trehalose contains the darkest contrast while pure lysozyme shows the lightest contrast. The second and third rows, both resulting in comparable images opposite to the first row, are based on specific wavelengths of lysozyme. The inhomogeneous character of the reference mixtures containing 50 and 80% lysozyme can be seen, although differences in contrast within these images seem relatively small.

Peak-Height Ratio Based Images

In Fig. 3 images of the 0, 10, 50, 80 and 100% lysozyme reference mixtures are given based on the peak-height ratio of 1,496/1,432 nm and 2,168/1,432 nm in order to enhance the spectral difference between lysozyme and trehalose. The images in both rows are visually almost identical, but the 2,168/1,432 nm ratio seems to result in a somewhat sharper image. The difference in contrast of the reference mixtures is clearly larger than in the single wavelength based images (cf. Fig. 2).

Correlation Coefficient Based Images

In Fig. 4 the images of the reference samples based on the correlation coefficient with lysozyme and with trehalose are given. In this approach information of the entire spectrum is used which could result in a higher sensitivity. In addition, the correlation coefficient can be calculated with lysozyme and trehalose independently. The difference in contrast between pure lysozyme and trehalose but also within the images of the reference mixtures based on the correlation coefficient in Fig. 4 is substantially larger than the difference in contrast in Figs. 1–3. As expected, the contrast of the images related to lysozyme (Fig. 4, upper row) is complementary to that of the images related to trehalose (Fig. 4, lower row).

Principal Component Analysis Based Images

In Fig. 5 the images of the reference samples are given based on the score value of the second principal component. The second principal component shows the strongest relation with the trehalose-lysozyme variation in comparison with the other principal components. From the images in Fig. 5 it can be concluded that the selected second principal component describes the lysozyme-trehalose variation. The loadings corresponding to the second principal component are given in Fig. 6. The images show more difference in contrast than

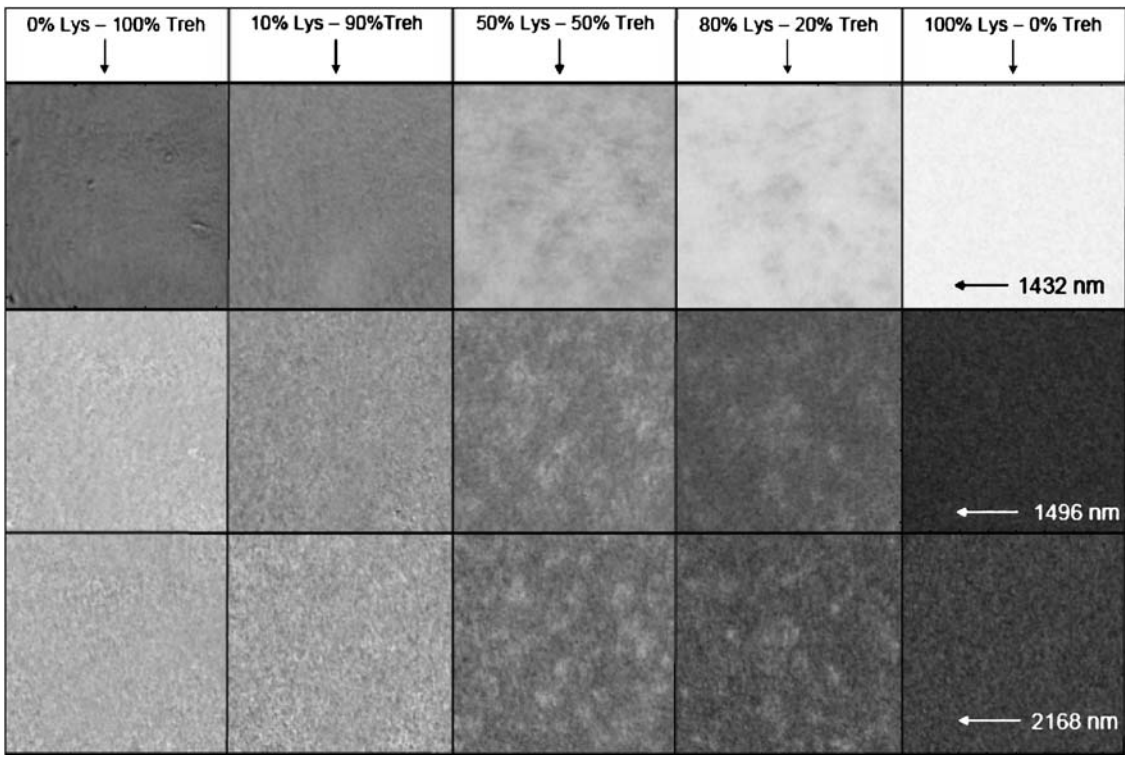


Fig. 2. Contrast images of lysozyme-trehalose (physical) mixtures containing 0, 10, 50, 80 and 100% lysozyme (160 × 160 pixels, pixel size = 40 μm) based on single wavelength intensity at 1,432, 1,496 and 2,168 nm of second derivative NIR spectra after normalization. The *higher* the peak intensity, the *darker* the contrast.

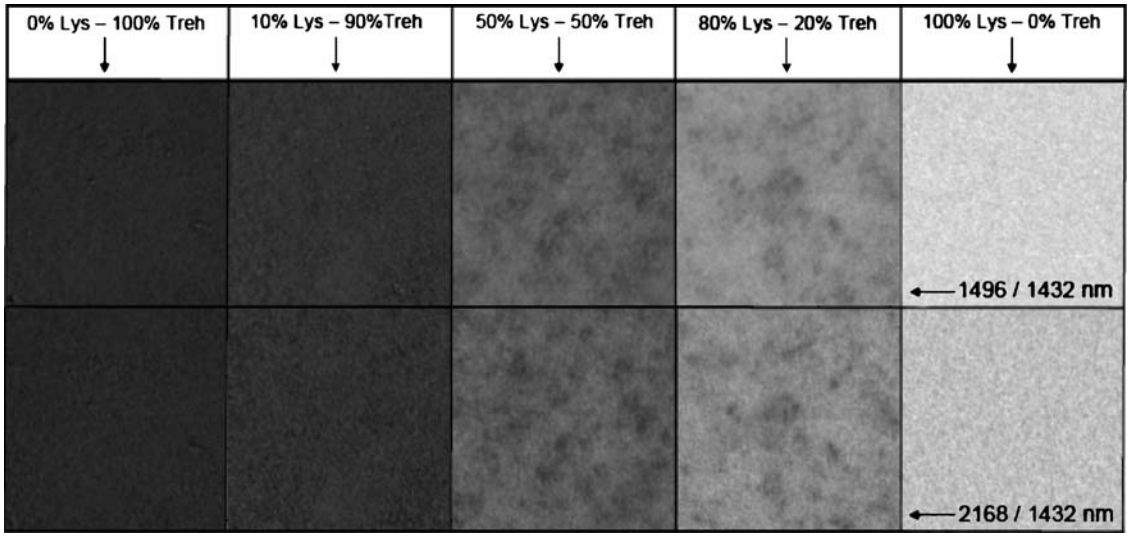


Fig. 3. Contrast images of lysozyme-trehalose (physical) mixtures containing 0, 10, 50, 80 and 100% lysozyme (160 × 160 pixels, pixel size 40 μm) based on peak-height ratio 1,496/1,432 and 2,168/1,432 nm of second derivative NIR spectra after normalization. A *dark contrast* corresponds to a low ratio.

the single wavelength and peak-height ratio based images, but somewhat less than the correlation coefficient based images. The loadings spectrum (Fig. 6) contains absorption bands characteristic for lysozyme and trehalose, which implies that those specific wavelengths are important for the variation in the contrast image. The shape of the loadings spectrum confirms that the second principal component is dependent on lysozyme and trehalose variation and that the second principal component can be used for describing sample homogeneity.

Methods for Homogeneity Assessment

It can be concluded that multivariate approaches (i.e., correlation coefficient and PCA analysis) result in a more distinct contrast than univariate approaches (single wavelength or peak ratio method). Correlation coefficient mapping and PLS regression were selected as data processing methods for further homogeneity investigations of dried samples. In order to assess the homogeneity, a (semi-) quantitative relation between the correlation coefficient and the concentration lysozyme and trehalose should be known. When this relation is known, the lysozyme and trehalose concentration are predicted independently by the pure component reference spectra. When using a PLS regression model, the resulting contrast of the images is proportional to the predicted lysozyme concentration, which is inversely proportional to the trehalose concentration. A PLS model is expected to be more robust than a correlation coefficient based method with respect to spectral variations that are not due to concentration differences, because the regression coefficients of the corresponding PLS factors are high for

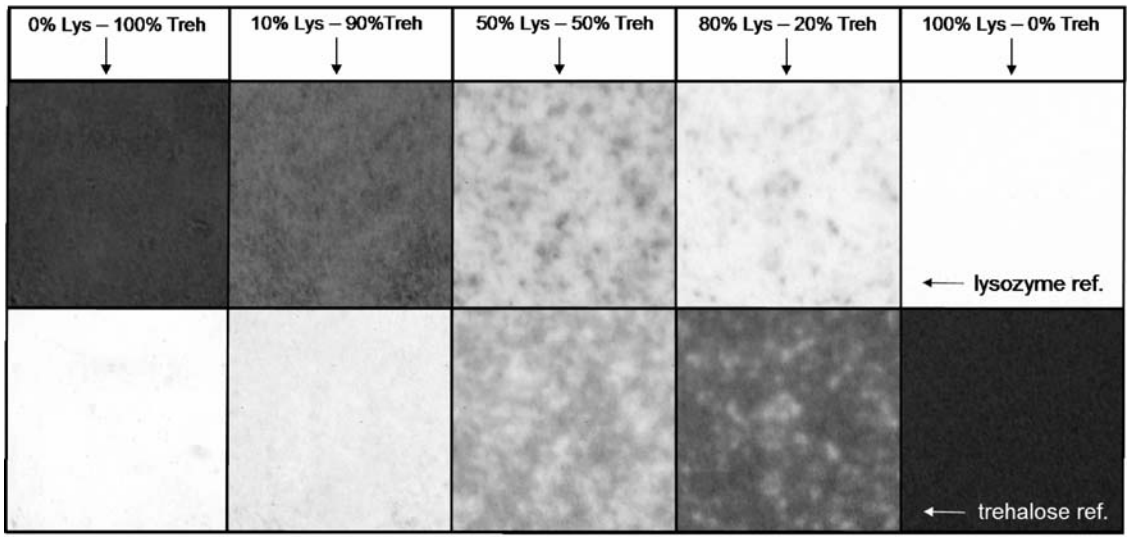


Fig. 4. Contrast images of lysozyme-trehalose (physical) mixtures containing 0, 10, 50, 80 and 100% lysozyme (160 × 160 pixels, pixel size 40 μm) based on correlation with pure lysozyme and pure trehalose, second derivative NIR spectra after normalization. A correlation of 1.000 corresponds to a *white contrast* image.

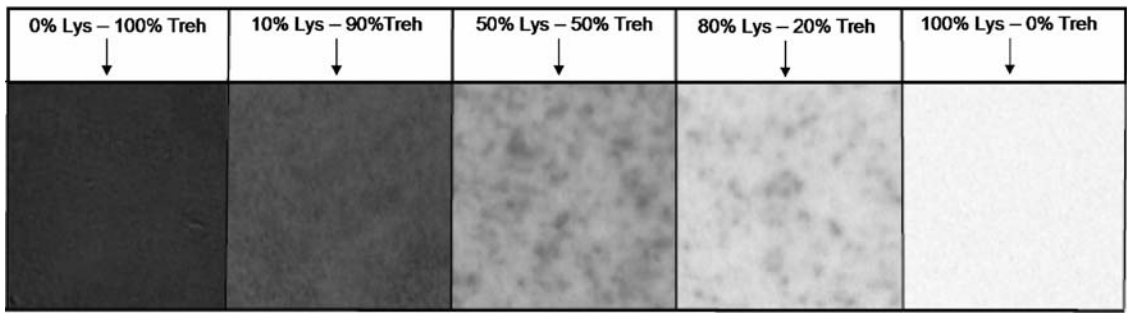


Fig. 5. Contrast images of lysozyme-trehalose mixtures containing 0, 10, 50, 80 and 100% lysozyme based on the score value of the second principal component, second derivative NIR spectra after normalization.

important wavelengths and low for less important or redundant wavelengths. In contrast to PLS regression, the correlation coefficient will decrease for every spectral variation regardless whether this variation is due to concentration differences of lysozyme and trehalose or not. It is expected that the precision of the predicted value using the correlation coefficient is higher for the concentrations around the pure components, while the precision is expected to be lower for mixtures. For PLS it is expected that the standard error of prediction is larger for the pure components (edge of the model range) compared to the mixtures.

Concentration versus Correlation Coefficient

When homogeneity is investigated based on the contrast of an image, it is necessary to know how the variation in contrast is related to the concentration of the component of interest. Theoretically the correlation coefficient must be related to the concentration of lysozyme and trehalose. In order to determine this relation, the correlation coefficient of the average spectra of the reference mixtures with pure lysozyme and pure trehalose was calculated and plotted against the corresponding concentration lysozyme. In Fig. 7 the relation between the correlation coefficient and the concentration of lysozyme or trehalose can be observed; the concentration of lysozyme and trehalose can be calculated from the correlation coefficient 'r' according to the following equations, which were obtained by second order least squares data fitting:

$$\textit{Lysozyme concentration}\ (\%) = 227.60(r_{Lys})^2 - 151.35(r_{Lys}) + 26.14 \quad (2)$$

$$\textit{Trehalose concentration}\ (\%) = 198.97(r_{Treh}) - 107.67(r_{Treh}) + 13.12 \quad (3)$$

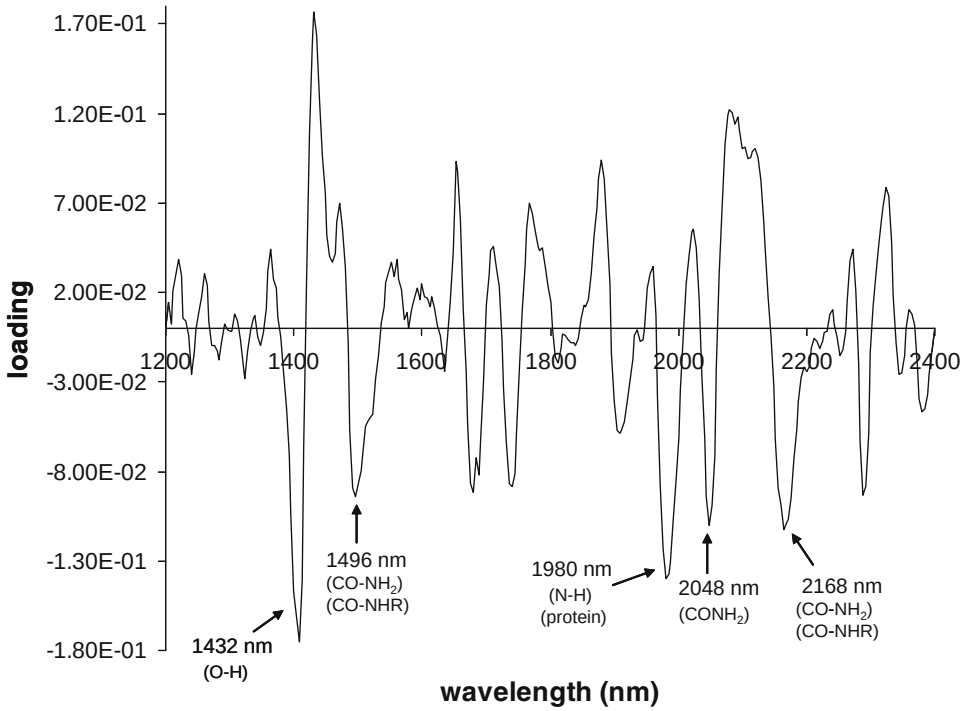


Fig. 6. Loadings corresponding to second principal component based on second derivative (SG 9/3) NIR spectra after normalization of lysozyme-trehalose reference samples.

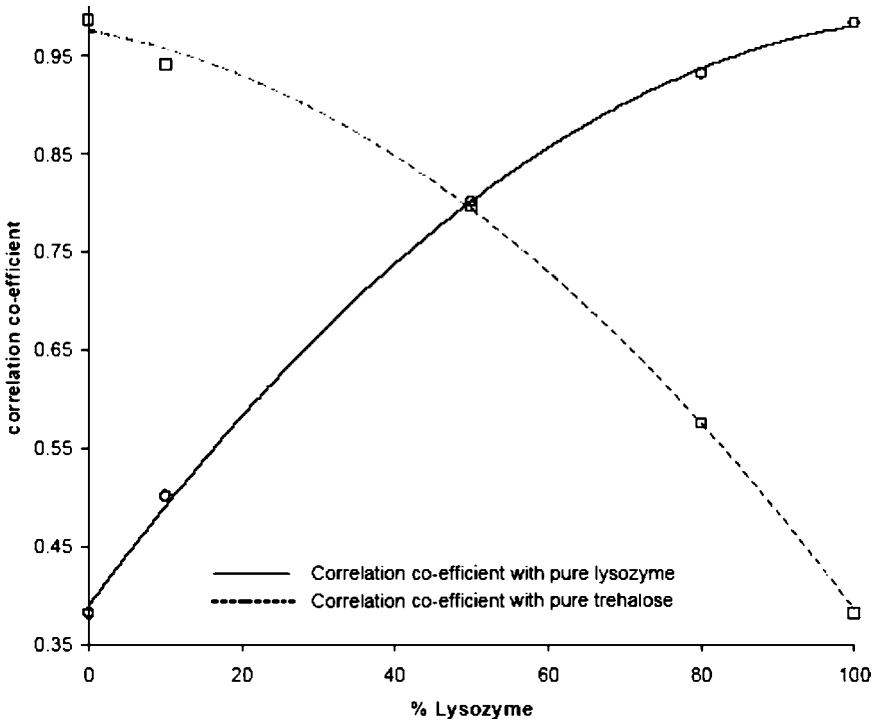


Fig. 7. Correlation coefficient of average image NIR spectra of reference samples (containing 0, 10, 50, 80 and 100% lysozyme) with pure lysozyme (*circles*) and pure trehalose (*squares*); second derivative NIR spectra, after normalization.

From the residuals of these relations between correlation coefficient and concentration as described in Fig. 7, the standard error for the prediction of the concentration lysozyme and trehalose was estimated at about 2% for both lysozyme and trehalose.

When the correlation coefficient with lysozyme is high, the correlation coefficient with trehalose is expected to be low, and vice versa, because the investigated mixtures are based on only these two components. In order to prove that areas within an image containing a lower correlation coefficient for lysozyme contain a higher correlation coefficient for trehalose, correlation coefficient images of the reference sample 50% lysozyme–50% trehalose with lysozyme and with trehalose were averaged into a new image. The average image should show a significantly lower standard deviation as compared to both individual images. Indeed, Fig. 8 shows that the images based on the correlation coefficients with lysozyme and trehalose show a much higher contrast and range in correlation coefficient than the average correlation image. It can therefore be assumed that the lysozyme and trehalose Eqs. (2) and (3), based on the correlation coefficient are specific with respect to lysozyme and trehalose.

PLS Regression

In PLS regression a number of significant factors, which individually describe a typical and independent variation between the spectra of the five reference mixtures, were used to predict the lysozyme (or trehalose) concentration directly from unknown samples. The average spectra of the reference mixtures were used to create a PLS model of two significant factors (first two PLS factors), which describes 98% of the

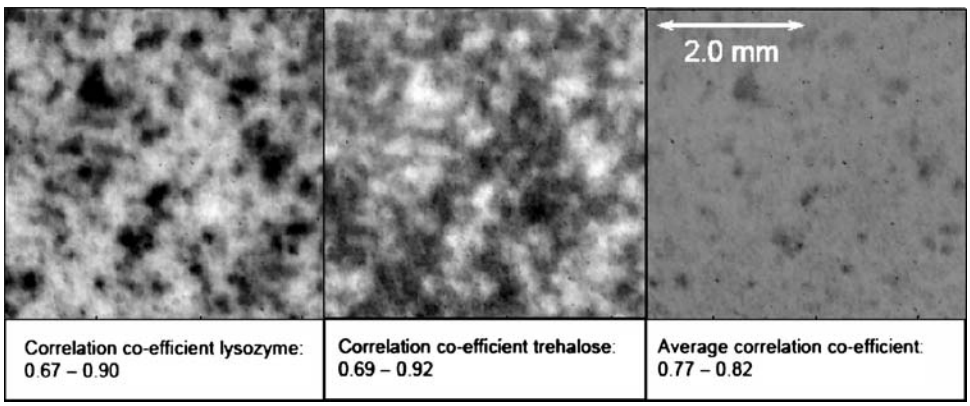


Fig. 8. Correlation coefficient image with lysozyme and trehalose and the resulting average image.

Table I. Concentration Range of Lysozyme and Trehalose for each Reference Sample Based on the Correlation Coefficient with Pure Lysozyme and Trehalose

	Lysozyme (%)	SD of correlation coefficient	Correlation coefficient with pure lysozyme			Difference[1] Min–max (%)
			Average	Min −3*SD	Max +3*SD	
1	100	0.0037	0.9835	0.9724	0.9946	6.6
2	80	0.0174	0.9316	0.8794	0.9838	28.5
3	50	0.0414	0.8028	0.6786	0.9270	53.2
4	10	0.0293	0.5001	0.4122	0.5880	13.4
5	0	0.0182	0.3809	0.3263	0.4355	2.4
	Trehalose (%)	SD of correlation coefficient	Correlation coefficient with pure Trehalose			Difference[2] Min–max (%)
			Average	Min −3*SD	Max +3*SD	
1	0	0.0164	0.3827	0.3335	0.4319	4.3
2	20	0.0410	0.5785	0.4555	0.7015	30.1
3	50	0.0384	0.8001	0.6849	0.9153	48.6
4	90	0.0108	0.9421	0.9097	0.9745	17.4
5	100	0.0062	0.9857	0.9671	1.0043	10.6

[1] Difference in min–max concentration based on Eq. (2).
[2] Difference in min–max concentration based on Eq. (3).

total spectral variance and 100% of the total variance of the reference values. Because the explained spectral variation of the PLS model is near 100%, nearly all spectral variation is important in the model. The total described variance of the reference values is 100% (rounded off), which implies that the model predicts nearly all the present variation of the reference values (exactly 100% = no calibration error). Based on full cross validation, the correlation coefficient between theoretical and predicted value equals 0.9981 and the estimated standard error of prediction is about 2%.

Spatial Resolution

Because of the penetration depth of NIR radiation, it is expected that each pixel contains a mixture of lysozyme and trehalose, unless very large (larger than the penetration depth) particles consisting of pure compounds are present. This effect is even more pronounced when powder samples are pressed into tablets. For homogeneity investigation the physical mixtures were recorded using both 10 and 40 μm resolutions. Both images showed heterogeneous distribution (authors' observation). However, the total sample area is directly related to the pixel size and for getting the best representative area and more accurate average reference spectra, 40 μm resolution images were selected for the reference mixtures. On the other hand, in order to enhance (spatial) resolution with respect to the detection of agglomerates of lysozyme or trehalose, images of the SCF dried and lyophilized samples were recorded using a magnification corresponding to a 10 μm pixel size (the highest possible spatial resolution) instead of a 40 μm pixel size.

Determination of Homogeneity

The homogeneity within an image is related to the variation with respect to the lysozyme and trehalose concentration. Within

Table IIa. Concentration Range of Lysozyme and Trehalose for SCF Dried and Freeze-Dried Samples Containing Lysozyme and Trehalose (1:1 (w/w)), Based on the Correlation Coefficient with Pure Lysozyme and Trehalose

Image	Sample	SD of correlation coefficient	Correlation coefficient with pure lysozyme			Calculated concentration[1] lysozyme			Difference Min–max (%)
			Average	Min −3*SD	Max +3*SD	Average (%)	Min (%)	Max (%)	
A	SCF dried	0.0075	0.7594	0.7369	0.7819	42.5	38.2	47.0	8.8
B	Freeze-dried	0.0100	0.7922	0.7621	0.8223	49.1	43.0	55.6	12.6
Image	Sample	SD of correlation coeffiecient	Correlation coefficient with pure trehalose			Calculated concentration[2] trehalose			Difference Min–Max (%)
			Average	Min −3*SD	Max +3*SD	Average (%)	Min (%)	Max (%)	
A	SCF dried	0.0071	0.7252	0.7039	0.7465	39.7	35.9	43.6	7.7
B	Freeze-dried	0.0099	0.7213	0.6916	0.7510	39.0	33.8	44.5	10.7

[1] Based on Eq. (2).
[2] Based on Eq. (3).

Table IIb. Concentration Range of Lysozyme and Trehalose for SCF Dried and Freeze-Dried Samples Containing Lysozyme and Trehalose (1:1 (w/w)), Based on the PLS Regression

		Predicted concentration Lysozyme			Predicted concentration Trehalose			Difference
Image	Sample	Average (%)	Min (%)	Max (%)	Average (%)	Min (%)	Max (%)	Min–Max (%)
A	SCF dried	58.0	54.0	62.4	42.0	37.6	46.0	8.4
B	Freeze-dried B	59.5	53.4	65.5	40.5	34.5	46.6	12.1

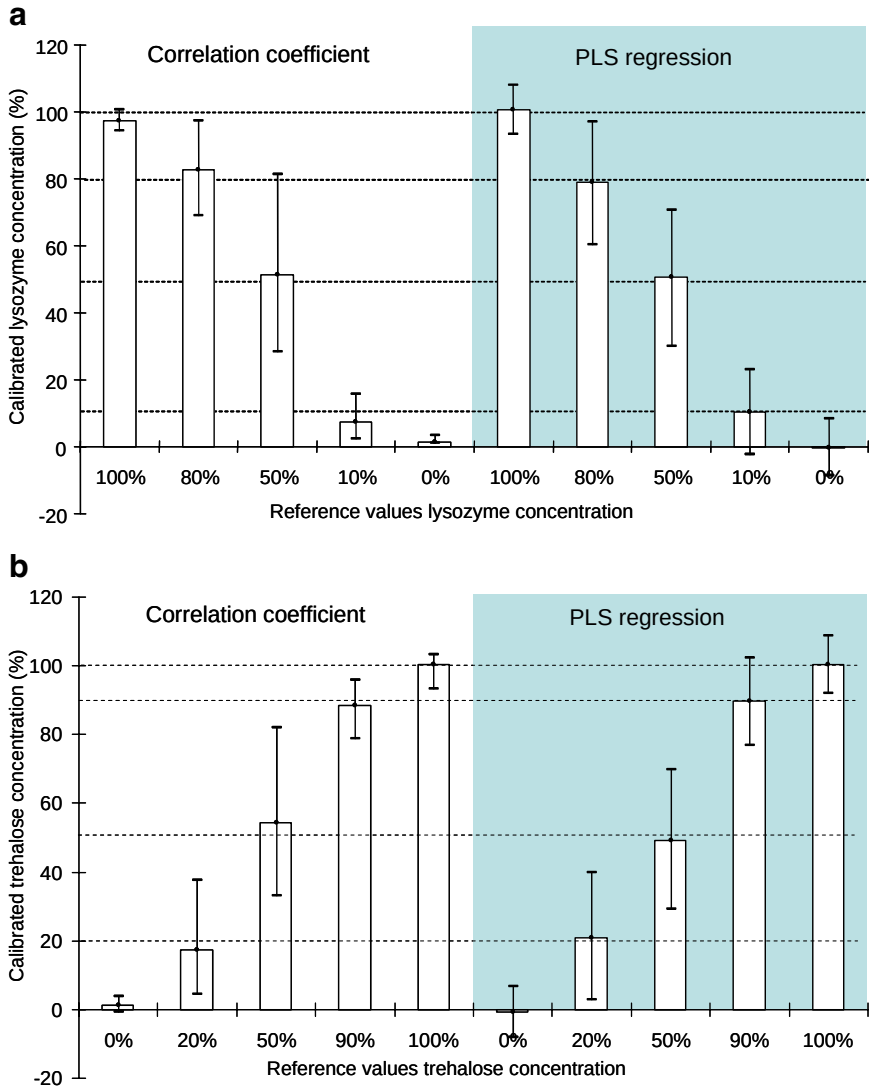
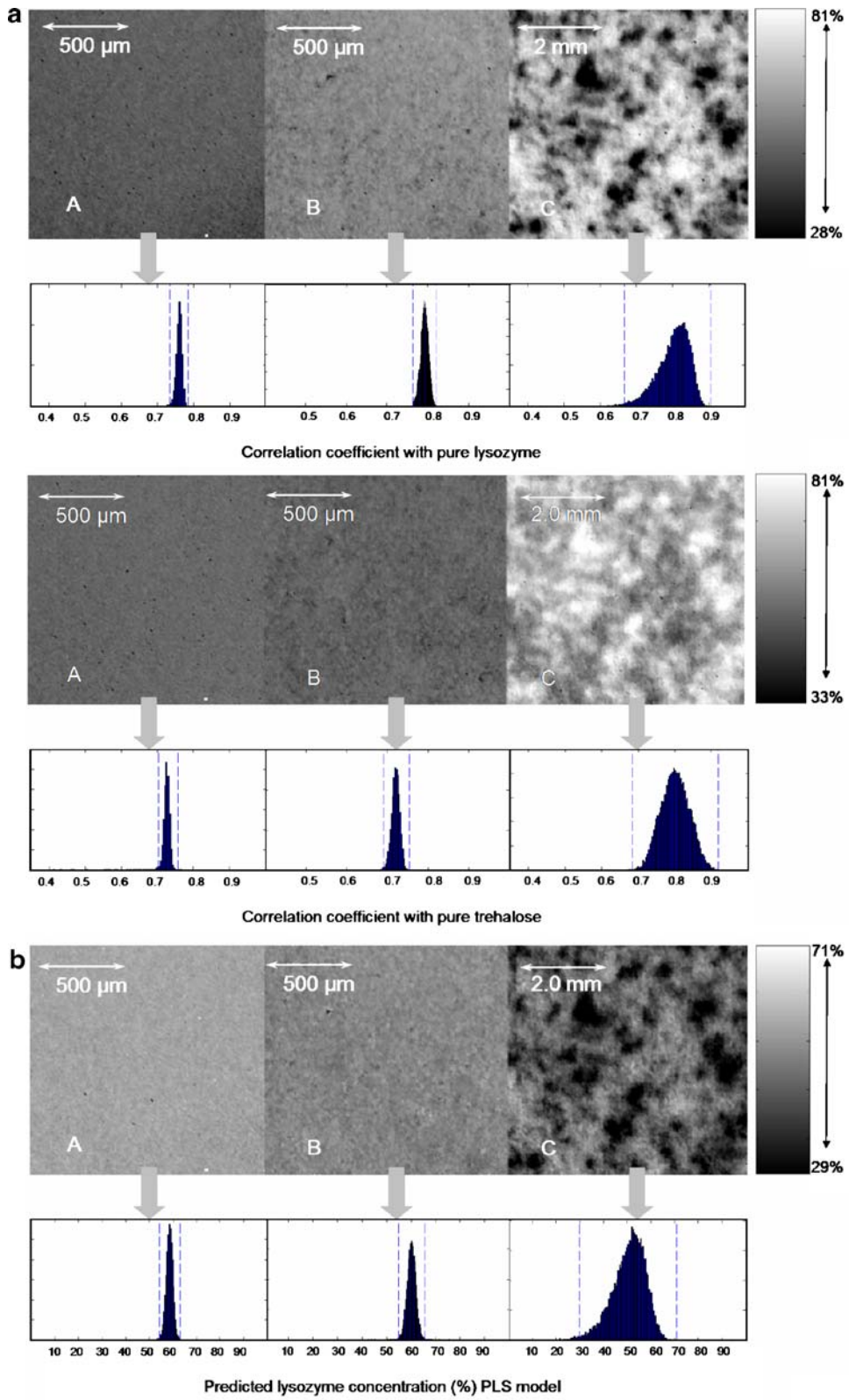


Fig. 9. (a) Calculated lysozyme concentration based on the correlation coefficient and PLS regression. *Block bars* represent average concentration and *error bars* represent min–max concentration range (±3 SD). (b) Calculated trehalose concentration based on the correlation coefficient and PLS regression. *Block bars* represent average concentration and *error bars* represent min–max concentration range (±3 SD).

Fig. 10. (a) Contrast images based on the correlation coefficient with pure lysozyme (*upper panel*) and pure trehalose (*lower panel*). Image A: SCF dried sample (160 × 160 pixels, pixel size 10 μm). Image B: freeze-dried sample (160 × 160 pixels, pixel size 10 μm). Image C: physical mixture of 50% lysozyme and 50% trehalose (160 × 160 pixels, pixel size 40 μm). Contrast bars related to the concentration of each component in the mixture are added on the right side of the images. Correlation coefficient distributions are shown below each image. (b) Contrast image based on predicted concentration lysozyme with PLS regression. Image A: SCF dried sample (160 × 160 pixels, pixel size 10 μm). Image B: freeze-dried sample (160 × 160 pixels, pixel size 10 μm). Image C: physical mixture of 50% lysozyme and 50% trehalose (160 × 160 pixels, pixel size 40 μm). A contrast bar representing the distribution of lysozyme concentration is added on the right side of the images, and the accompanying distributions of the predicted concentrations below each image.


a
500 µm
500 µm
2 mm
A
B
C
81%
28%
0.4 0.5 0.6 0.7 0.8 0.9
0.5 0.6 0.7 0.8 0.9
0.4 0.5 0.6 0.7 0.8 0.9
Correlation coefficient with pure lysozyme
500 µm
500 µm
2.0 mm
A
B
C
81%
33%
0.4 0.5 0.6 0.7 0.8 0.9
0.5 0.6 0.7 0.8 0.9
0.4 0.5 0.6 0.7 0.8 0.9
Correlation coefficient with pure trehalose
b
500 µm
500 µm
2.0 mm
A
B
C
71%
29%
10 20 30 40 50 60 70 80 90
10 20 30 40 50 60 70 80 90
10 20 30 40 50 60 70 80 90
Predicted lysozyme concentration (%) PLS model

one image, the distribution of the correlation coefficient or predicted concentration from the PLS model was shown to be a measure for homogeneity. Furthermore, the differences between calculated concentrations are also used for comparison of homogeneity between references and samples (as shown in the last column in Tables I, IIa, IIb). The standard deviation (SD) of the distribution of the concentration was used to assess homogeneity. The SD of a pure component image was considered to be representative of a homogeneous sample. By using a minimum and maximum value of three times SD, nearly 100% of the concentration range within an image is covered. In Table I the correlation coefficients and their standard deviation are shown for lysozyme and trehalose. Lysozyme and trehalose concentrations derived from the correlation coefficient, based on Eqs. (2) and (3), and the PLS model are shown in Fig. 9a and b.

From Table I and Fig. 9a and b, it can be concluded that the reference mixture containing 50% lysozyme and 50% trehalose is, as expected, most inhomogeneous. The concentration range of lysozyme from 28.2 to 81.4% covers a concentration difference of nearly 50% based on the correlation coefficient and about 40% based on PLS regression. Pure lysozyme and trehalose show a concentration difference of 6.6 and 2.4%, respectively, based on the correlation coefficient method, due to spectral variance, which is not related to concentration differences but to spectral noise and probably physical differences within a sample, like surface and density differences. The PLS regression model shows a wider concentration range for the pure compounds of about 15%. Because the correlation coefficient cannot become >1, it is expected for the correlation coefficient to show a somewhat smaller concentration range compared to the normal distributed PLS prediction.

Application of NIR for Studying Sample Homogeneity

In our previous work (7) we indicated possible inhomogeneity in SCF dried protein-sugar formulations. Therefore, using NIR imaging we compared the homogeneity of a SCF dried formulation with that of a freeze-dried formulation, both containing equal fractions of lysozyme and trehalose.

Images of the SCF dried sample and the freeze dried sample were recorded using a resolution corresponding to a 10 μm pixel size. The contrast of the images based on the correlation coefficient with lysozyme and trehalose (Fig. 10a) and PLS regression (Fig. 10b) was compared to the reference sample, a physical mixture containing 50% lysozyme and 50% trehalose (images C), recorded with a resolution corresponding to a 40 μm pixel size. The reference sample shows heterogeneous images (Fig. 10, images C) with clearly visible areas where the local lysozyme concentration significantly deviates from the average lysozyme concentration, which is in line with the above results. The concentration range of lysozyme in the physical mixture covers 28 to 81% while the trehalose concentration varies complementarily and ranges from 33 to 81% (Fig. 10a). Based on the PLS model, the lysozyme concentration range of the reference sample is somewhat smaller and lies between 29 and 70%.

The images of the SCF dried sample (images A in Fig. 10) do not show visual marks of inhomogeneity based on correlation coefficient analysis with lysozyme and trehalose (Fig. 10a) or PLS regression analysis (Fig. 10b). Accordingly, the correlation coefficient distributions of the SCF dried sample images are much narrower than the distribution within the reference (physical mixture). Based on the correlation coefficient, calculated concentration ranges in the SCF dried sample are 38–47 and 36–44%, for lysozyme and trehalose, respectively (Table IIa). So, the calculated differences between the maximum and minimum concentrations are 8–9%, which is comparable to the calculated differences for the pure compounds (cf. Table I) and much smaller than the differences of about 50% for the reference mixture (Fig. 11). Also based on PLS prediction, a similarly small lysozyme concentration difference (8.4%) is observed for the SCF dried sample as compared to the physical mixture (Table IIb, Fig. 11). Altogether, these data point to a homogeneous SCF dried sample at the resolution level of the NIR imaging method (10 μm pixel size).

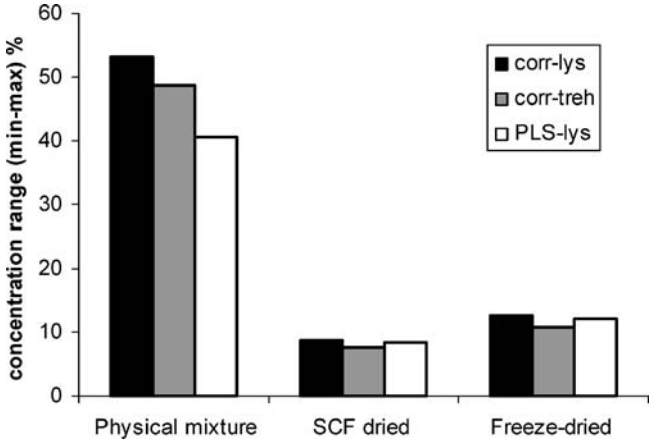


Fig. 11. Homogeneity based on the calculated concentration range (differences in concentration minimum and maximum) of SCF dried and freeze-dried samples compared to reference sample (physical mixture of 50% lysozyme and 50% trehalose). Shown concentration ranges are calculated based on correlation coefficient with lysozyme (*black bars*) and trehalose (*grey bars*), and on PLS (*white bars*).

The freeze-dried sample (Fig. 10, images B) is visually homogeneous when compared to the physical mixture (images C) but slightly less homogeneous compared to the SCF dried sample (images A), as reflected by the distributions of the correlation coefficients (Fig. 10a) and those of the lysozyme concentration (Fig. 10b and Fig. 11). The average concentrations and concentration ranges based on the correlation coefficient and the PLS model are given in Tables IIa and IIb.

The average concentrations of lysozyme and trehalose of the SCF dried sample and the freeze dried sample deviate from expected 50%. Regarding the precision of the correlation coefficient model and the PLS model individually a smaller deviation between the two models would be expected. The PLS model should provide the most accurate prediction at the 50% concentration level compared to the correlation coefficient model. The correlation coefficient based model could be sensitive to the other variations than lysozyme and trehalose concentration especially considering the fact that the calibration samples were prepared from the mixtures of pure materials and not from the SCF or the

freeze dried sample. Therefore, an extra spectral variation in the SCF and freeze dried sample could cause this concentration deviation. However, this was not considered as critical in order to assess concentration differences within the samples for the homogeneity investigation.

Information on homogeneity of SCF dried powders is very important for further development of SCF processes for stabilizing proteins. In our previous work (7) we suggested inhomogeneity of the sugar matrix as a possible reason for protein instability in some of the SCF dried powders when compared to freeze-dried formulations. The data of the present study indicate that the problem of observed protein instability in SCF drying lies not in inhomogeneous distribution of protein and sugar but rather in other formulation and/or processing parameters, which will be addressed in follow-up studies.

The application of NIR imaging for homogeneity assessment could be broader than discussed in this investigation. As a fast, non-invasive technique, NIR imaging can be used as process analytical technology (PAT) for dried protein-containing dosage forms. Furthermore, NIR imaging could be useful for monitoring changes in heterogeneity during (accelerated) storage of dried formulations. Also, the heterogeneity of freeze-dried cakes as a function of height could be studied after physical separation of cake layers (e.g., a glaze on top of a cake *versus* lower layers).

CONCLUSIONS

In this work we investigated the possibility to apply NIR imaging for studying the homogeneity of protein-sugar mixtures. Among the imaging approaches studied, correlation coefficient mapping provides the most discriminating spectral images regarding the lysozyme and trehalose concentration. In addition, there is a clear relation between correlation coefficient and concentration, which makes the correlation coefficient based images suitable for quantification of homogeneity. Besides the correlation coefficient, the PLS regression model is also suitable for quantitative interpretation of homogeneity. The PLS regression model is somewhat more accurate than the correlation coefficient method, whereas at the concentration limits, e.g., 100% lysozyme or 0% lysozyme, the correlation coefficient model is more accurate. Both models clearly showed homogeneity of a SCF dried lysozyme-trehalose mixture as compared to a freeze-dried mixture of the same composition. So, NIR imaging in combination with the two imaging approaches is a suitable technique for studying homogeneity of dried protein formulations.

ACKNOWLEDGMENTS

This research was supported in part by the Technology Foundation STW, applied science division of NWO and the technology program of the Dutch Ministry of Economic Affairs.

REFERENCES

1. J. F. Carpenter and M. C. Manning. *Rational Design of Stable Protein Formulations, Theory and Practice*. Kluwer Academic/Plenum Publisher, 2002.
2. M. J. Pikal. Freeze-drying of proteins. Part II: Formulation selection. *BioPharm* **3**:26–30 (1990).
3. H. R. Costantino and M. J. Pikal. *Lyophilization of Biopharmaceuticals*, American Association of Pharmaceutical Sciences, Arlington, 2004.
4. N. Jovanovic, A. Bouchard, G. W. Hofland, G. J. Witkamp, D. J. A. Crommelin, and W. Jiskoot. Stabilization of proteins in dry powder formulations using supercritical fluid technology. *Pharm. Res.* **21**:1955–1969 (2004).
5. A. Dong, S. J. Prestrelski, S. D. Allison, and J. F. Carpenter. Infrared spectroscopic studies of lyophilization- and temperature-induced protein aggregation. *J. Pharm. Sci.* **84**:415–424 (1995).
6. S. J. Prestrelski, T. Arakawa, and J. F. Carpenter. Separation of freezing- and drying-induced denaturation of lyophilized proteins using stress-specific stabilization. II. Structural studies using infrared spectroscopy. *Arch. Biochem. Biophys.* **303**:465–473 (1993).
7. N. Jovanovic, A. Bouchard, G. W. Hofland, G. J. Witkamp, D. J. Crommelin, and W. Jiskoot. Distinct effects of sucrose and trehalose on protein stability during supercritical fluid drying and freeze-drying. *Eur. J. Pharm. Sci.* in press: (2006).
8. S. T. Tzannis and S. J. Prestrelski. Activity-stability considerations of trypsinogen during spray drying: effects of sucrose. *J. Pharm. Sci.* **88**:351–359 (1999).
9. K. Izutsu, S. Yoshioka, and T. Terao. Decreased protein-stabilizing effects of cryoprotectants due to crystallization. *Pharm. Res.* **10**:1232–1237 (1993).
10. T. W. Randolph. Phase separation of excipients during lyophilization: effects on protein stability. *J. Pharm. Sci.* **86**:1198–1203 (1997).
11. M. C. Heller, J. F. Carpenter, and T. W. Randolph. Manipulation of lyophilization-induced phase separation: implications for pharmaceutical proteins. *Biotechnol. Prog.* **13**:590–596 (1997).
12. M. C. Heller, J. F. Carpenter, and T. W. Randolph. Conformational stability of lyophilized PEGylated proteins in a phase-separating system. *J. Pharm. Sci.* **88**:58–64 (1999).
13. W. Q. Sun and P. Davidson. Protein inactivation in amorphous sucrose and trehalose matrices: effects of phase separation and crystallization. *Biochim. Biophys. Acta* **1425**:235–244 (1998).
14. R. C. Lyon, D. S. Lester, E. N. Lewis, E. Lee, L. X. Yu, E. H. Jefferson, and A. S. Hussain. Near-infrared spectral imaging for quality assurance of pharmaceutical products: analysis of tablets to assess powder blend homogeneity. *AAPS PharmSciTech* **3**:E17, (2002).
15. A. S. El-Hagrasy, H. R. Morris, F. D'Amico, R. A. Lodder, and J. K. Drennen 3rd. Near-infrared spectroscopy and imaging for the monitoring of powder blend homogeneity. *J. Pharm. Sci.* **90**:1298–1307 (2001).
16. F. C. Clarke, S. V. Hammond, R. D. Jee, and A. C. Moffat. Determination of the information depth and sample size for the analysis of pharmaceutical materials using reflectance near-IR microscopy. *Appl. Spectrosc.* **56**:1475–1483 (2002).
17. A. Saleki-Gerhardt and G. Zografi. Non-isothermal and isothermal crystallization of sucrose from the amorphous state. *Pharm. Res.* **11**:1166–1173 (1994).
18. L. Eriksson, E. Johansson, N. Kettaneh-Wold, and S. Wold. *Multi- and Megavariate Data Analysis, Principles and Applications*. Umetrics AB, 2001.
19. D. L. Massart. *Datahandling in Science and Technology, Chemometrics: A Text Book*. Elsevier, 1988.
20. D. A. Burns and E. W. Ciurczak. *Handbook of Near-Infrared Analysis*, 2001.
21. http://www.ft-nir.com/Appendix/nir_absorption_table.htm.