

Determination of Protein Crosslinking with Bootstrap Pattern Selection and Near-Infrared Spectrophotometry

Jennifer Moses, Robert G. Buice, Jr., and Robert A. Lodder*

Division of Pharmaceutical Sciences, College of Pharmacy, University of Kentucky, Lexington, KY 40536-0082.

*Author to whom correspondence should be addressed.

The migration of formaldehyde from a polyethylene glycol (PEG) fill into the gelatin shell of a soft elastic capsule was observed using near-infrared (NIR) spectrophotometry. Bootstrap pattern selection (BPS) modeling was employed to analyze the spectra of the empty capsules. Suitable correlation was confirmed ($r^2 = 0.988$) when actual concentrations of formaldehyde in the PEG fill of the capsules were regressed against the values from NIR spectra of the emptied and washed capsules.

INTRODUCTION

Gelatin capsules are among the most versatile pharmaceutical oral delivery system, and one in which drug powders, solutions, or suspensions may be filled. Additional applications that employ the hard gelatin capsule (HGC), soft elastic gelatin capsule (SEGC), or gelatin coating include controlled release technologies, cosmetic formulations, and nutritional supplements. The beneficial properties of the gelatin capsule include its hardy yet pliant backbone, polished appearance, capability to hold dyes, and ease of swallowing (1).

The propensity of the gelatin molecule toward formaldehyde-induced derivatization and crosslinking through its amino and guanidino functionalities has been well described (2-6). The development of crosslinks between these tertiary functionalities of the polypeptide leads to an elevation in the molecular weight of gelatin and more specifically, a deleterious effect on the dissolution of the drug that it encapsulates (7). The likely effects of formaldehyde crosslinking in gelatin have been examined because of apparent dissolution problems in *in vitro* analyses of hard and soft gelatin capsules (1,8). It has been demonstrated that SEGCs containing 20 or 80 ppm formaldehyde in a lipophilic fill exhibit decreased dissolution rates of acetaminophen in contrast to fresh, unadulterated SEGCs (9). Utilizing HGCs exposed to atmospheric formaldehyde concentrations of 150 ppb for as few as two hours, Gold et al. observed partial insolubilization of the gelatin shell and diminished dissolution rates of amoxicillin when compared to unstressed capsules (10). These crosslinking events have been recognized to effect the development of a thin, water-insoluble membrane (pellicle) around the gelatin

capsule during dissolution testing (8). The pellicle behaves as a barrier that controls the diffusion of the drug. Recently, attention has been given to the origin of the low molecular weight aldehydes involved in gelatin capsule insolubilization. It has been revealed, for example, that corn starch, a frequent drug excipient and fill material in HGCs, may embody low levels of hexamethylenetetramine stabilizer (1, 11), which decomposes in damp conditions to produce ammonia and formaldehyde. Polyethylene glycols, commonly used as bases for hydrophobic drugs in soft elastic gelatin capsules (SEGCs), unleash low molecular weight aldehydes through free radical reactions and upon exposure to aerobic environments (11,12).

Recently, NIR spectrophotometry has been successfully employed for identification of water uptake (13,14) and extent of crosslinking in HGCs (10,13). In the latter investigation, evacuated gelatin capsules (lacking excipients or drug powders) were exposed to formaldehyde vapor. The capsules were then charged with fresh amoxicillin and their dissolution performance was determined *in vitro* as well as predicted successfully using NIR spectrophotometry (10,13). It is important to demonstrate, however, that a nondestructive and noninvasive method is able to detect formaldehyde that emanates from the encapsulated excipient and migrates into the gelatin shell, with the potential to render the gelatin insoluble.

The current research aggregates the analytical strength of NIR spectrophotometry and a new BEST regression with bootstrap model selection (15) in much the same way principal component regression has been used (16). These two techniques are applied to predict, in empty SEGCs, the concentration of formaldehyde in a polyethylene glycol base originally held inside the same capsules.

THEORY

Near-IR spectra are used in this research qualitatively and quantitatively in the analysis of protein crosslinks. Spectra are represented as elements of a matrix T , with one row for each of n samples and one column for each of d wavenumbers. Each sample point is translated from the origin in hyperspace in each dimension d by an amount that corresponds to the magnitude of the signal observed at each wavenumber. Under this scheme, similar samples give rise to similar spectra that project into similar regions of

hyperspace.

Let α be a subset of $\{1, \dots, p\}$ of size p_α and $t_{i\alpha}$ (or β_α) be the subvectors of t_{ij} (or β) containing the components of t_{ij} (or β) indexed by the integers in α . Then a model corresponding to α , designated model α for clarity, is

$$\mu_{i\alpha} = E(y_i | t_{ij}) = t'_{i\alpha} \beta_\alpha, \text{ var}(y_i | t_{ij}) = \sigma^2, \quad i = 1, \dots, n. \quad \text{Eq 1}$$

For a given α , model α is not necessarily a correct model in the sense that $E(y_i | t_{ij})$ is actually not always equal to $t'_{i\alpha} \beta_\alpha$. If β_α contains all nonzero components of β , then $t'_{i\alpha} \beta_\alpha = t'_{i\alpha} \beta$ for any t_{ij} and model (2) is called a correct model. There may also be additional correct models.

If α_0 is the subset matching the correct model with the smallest size; that is, β_{α_0} contains exactly all nonzero components of β , Then,

$$\liminf_{n \rightarrow \infty} \Delta_n(\alpha) > 0 \quad \text{for any incorrect model } \alpha, \quad \text{Eq 2}$$

and model α_0 is optimal in the sense that it minimizes $L_n(\alpha)$ over $\alpha \in A$ for sufficiently large n ; that is,

$$\lim_{n \rightarrow \infty} P\{L_n(\alpha_0) = \min_{\alpha \in A} L_n(\alpha)\} = 1. \quad \text{Eq 3}$$

Because $L_n(\alpha)$ comprises the unknown parameter β , the optimal α_0 must be estimated. Selecting a model is the same as finding an estimate of α_0 .

Let $\hat{\alpha}$ be the estimate of α based on a model selection procedure. Then the model selection procedure is defined as consistent if

$$\lim_{n \rightarrow \infty} P\{\hat{\alpha} = \alpha_0\} = 1. \quad \text{Eq 4}$$

The regression algorithm used analyzes distributions of spectra in hyperspace in an effort to create consistent, correct models. Following the initialization of the matrix \mathbf{T} , the algorithm calibrates similarly to the parallel assimilation modeling procedure in ref. 17. Three simple equations define the distance and direction relationships of the calibration samples, the test sample(s), and the calibration probability distribution:

$$S_{02} = \left(\sum_{j=1}^d (x_j - c_j)^2 \right)^{1/2} \quad \text{Eq 5}$$

$$S_{(OR)_i} = \left(\sum_{j=1}^d (b_{ij} - c_j)^2 \right)^{1/2} \quad \text{Eq 6}$$

$$S_{(2R)_i} = \left(\sum_{j=1}^d (b_{ij} - x_j)^2 \right)^{1/2} \quad \text{Eq 7}$$

An advantage of the BPS algorithm over the BEST assimilation modeling procedure is realized during the analysis of actual samples after calibration, when a series of equations that formerly had to be iterated can be replaced with simply

$$S_{(p)_i} = S_{(OR)_i}^2 - \left(\{ S_{(OR)_i}^4 + 2(S_{(OR)_i})^2(S_{(2R)_i})^2 - S_{(2R)_i}^4 + 2(S_{(OR)_i})^2(S_{(02)})^2 + 2(S_{(2R)_i})^2(S_{(02)})^2 - S_{(02)}^4 \} / 4(S_{(02)})^2 \right)^{1/2} \quad \text{Eq 8}$$

Standard deviations of $S_{(p)_i}$ are then used for qualitative and quantitative analysis of the plaque samples. The direction of S_{02} contains the qualitative information and the SDs of $S_{(p)_i}$ along S_{02} provides quantitative information for sample regression and prediction.

EXPERIMENTAL

Materials. Size 8, clear oblong capsules with 0.5 em twist-off tips were provided by Banner Pharmacaps, Inc. (Chatsworth, CA). Low-aldehyde polyethylene glycol (PEG) 400 was acquired from Union Carbide (Danbury, CT). Formaldehyde, 37% w/w in water, was obtained from Aldrich (St. Paul, MN).

Instrumentation. Spectral data were collected by a spectrophotometer assembled from a concave holographic grating and lead sulfide detector. Absorbance values were obtained as $\log(1/\text{reflectance})$ data from capsules from 1100 to 2500 nm. The data were collected and analyzed with an Intel Pentium computer running Windows software. Analytical software was written in Mathematica (Wolfram Research, Inc., Champaign, IL).

Procedure. Three capsules were each injected with one of five solutions. The solutions were PEG 400 (neat), 0.05 vol% formaldehyde in PEG 400, 0.10 vol% formaldehyde in PEG 400, 0.20 vol% formaldehyde in PEG 400, and 0.40 vol% formaldehyde in PEG 400. Injection of 0.5 ml of solution was executed by penetrating the capsule at the apex of the twist-off tube using a 1.0 ml syringe furnished with a 27 gauge needle. The capsules were cleaned at the point of injection to guarantee that the solution would not migrate to the outer surfaces of the capsule. The capsules were positioned with the apex up inside a sealed glass vessel for 48 h. Afterward, the tips of the SEGCs were excised, with care taken not to permit the solution inside to reach the outer capsule surfaces. Using a miniature pipet, the solutions inside the SEGCs were extracted, and the interior surfaces of the capsules were rinsed twice with new PEG 400 and twice with diethyl ether. The capsules were subsequently deposited in a nitrogen atmosphere for twenty four hours. Capsules were scanned three times with the lanced end downward in a 90° conical

reflector machined from aluminum (17). A steel wire brace was used to achieve vertical capsule positioning within the cone (10).

RESULTS AND DISCUSSION

Figure 1 presents the NIR spectra for all six experimental groups of evacuated SEGCs. Three reflectance values are entered at each wavelength because three NIR scans were collected from each sample. Spectral averaging, which preserved only the average reflectance value at each wavelength, was executed on the data. The spectra were subsequently smoothed with a cubic spline algorithm, after which spectral scattering effects arising from sample inhomogeneity (variation in capsule wall thickness and composition) were eliminated by multiplicative scatter correction. Nonuniformity of samples precluded simple analysis techniques from not only defining the spectral signal of an individual capsule, but also from discriminating

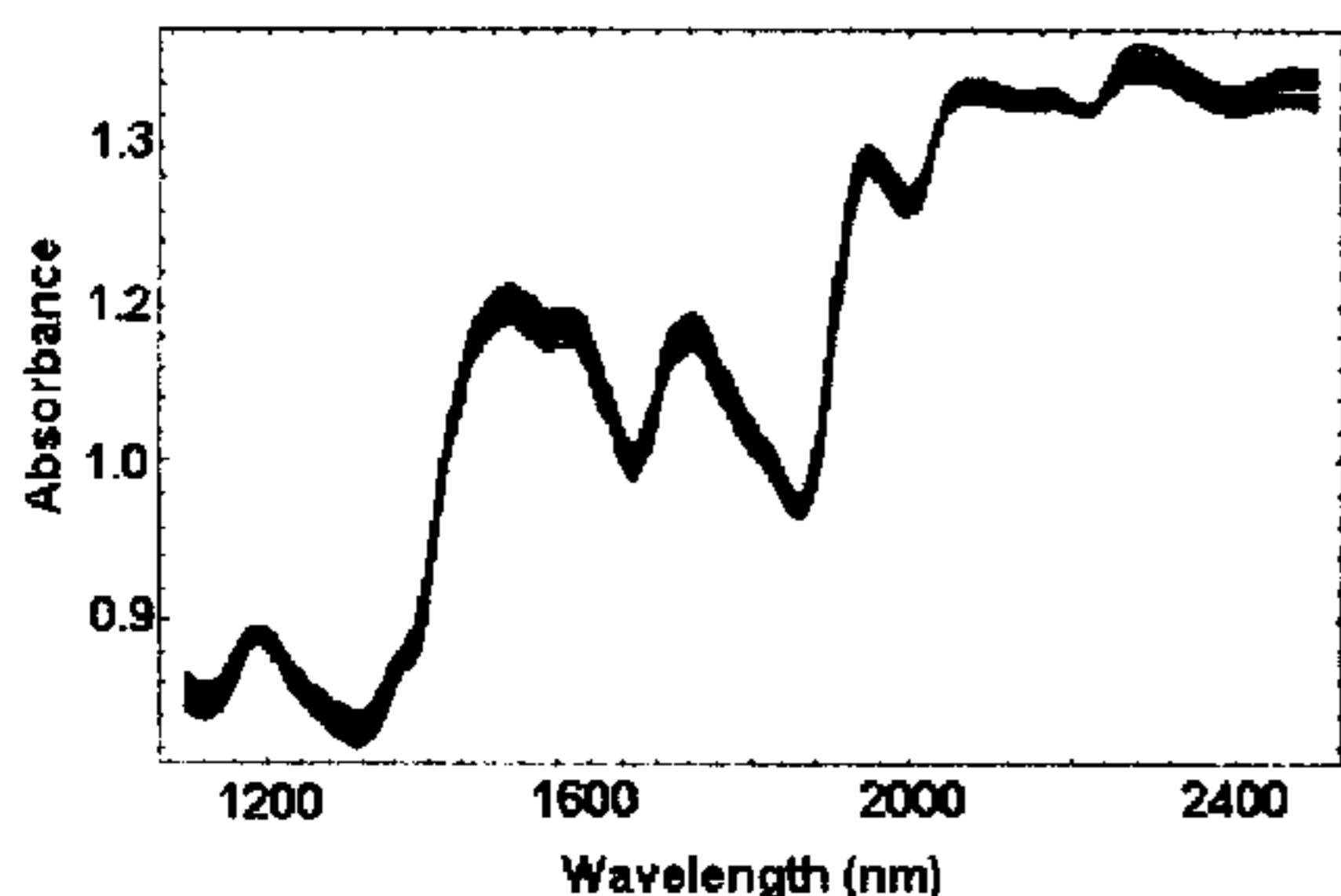


Figure 1 NIR spectra of empty soft gelatin capsules, which had been filled with polyethylene glycol +/- formaldehyde solution.

Accuracy and Precision

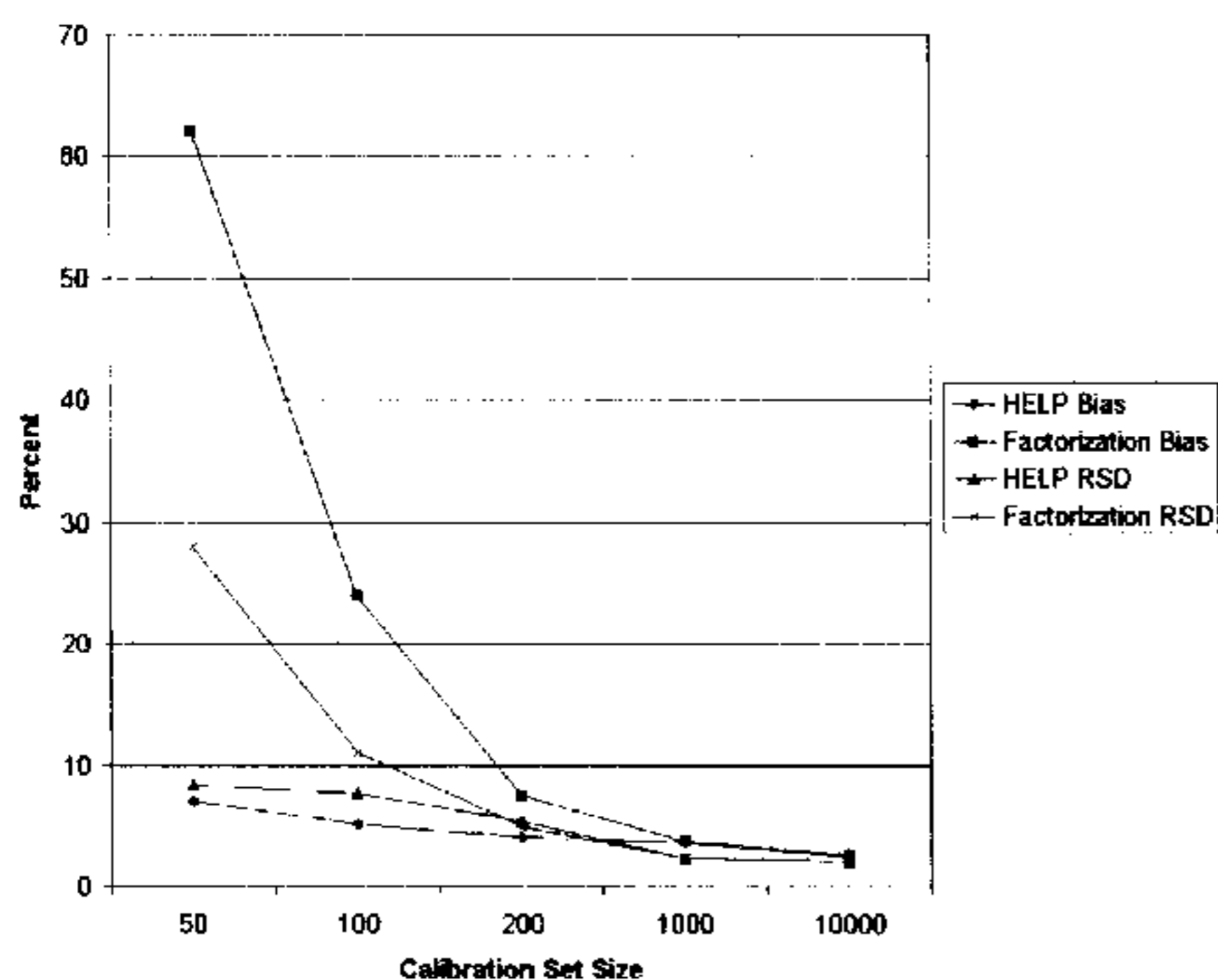


Figure 2 Accuracy and precision results as a function of number of calibration samples. \blacklozenge =BPS bias, \blacksquare =factorization bias, \blacktriangle =BPS RSD, \times =factorization RSD. BPS has more accuracy and precision than matrix factorization.

its NIR spectra from that of other capsules. Therefore, regression with bootstrap model selection was used to analyze the smoothed, scatter-corrected capsule spectra. Cross-validation was employed on the entire group of capsules. Figure 2 is a graph of the bias and RSD (using the ASTM E456 definitions) test results of the BPS algorithm on $N(0,1)$ data as a function of number of calibration set size. Figure 3 portrays the clear linear correlation ($r^2=0.988$) of actual concentration of formaldehyde (solution) in the PEG solutions, initially encapsulated by the SEGCs, to concentrations calculated using the NIR spectra of the same

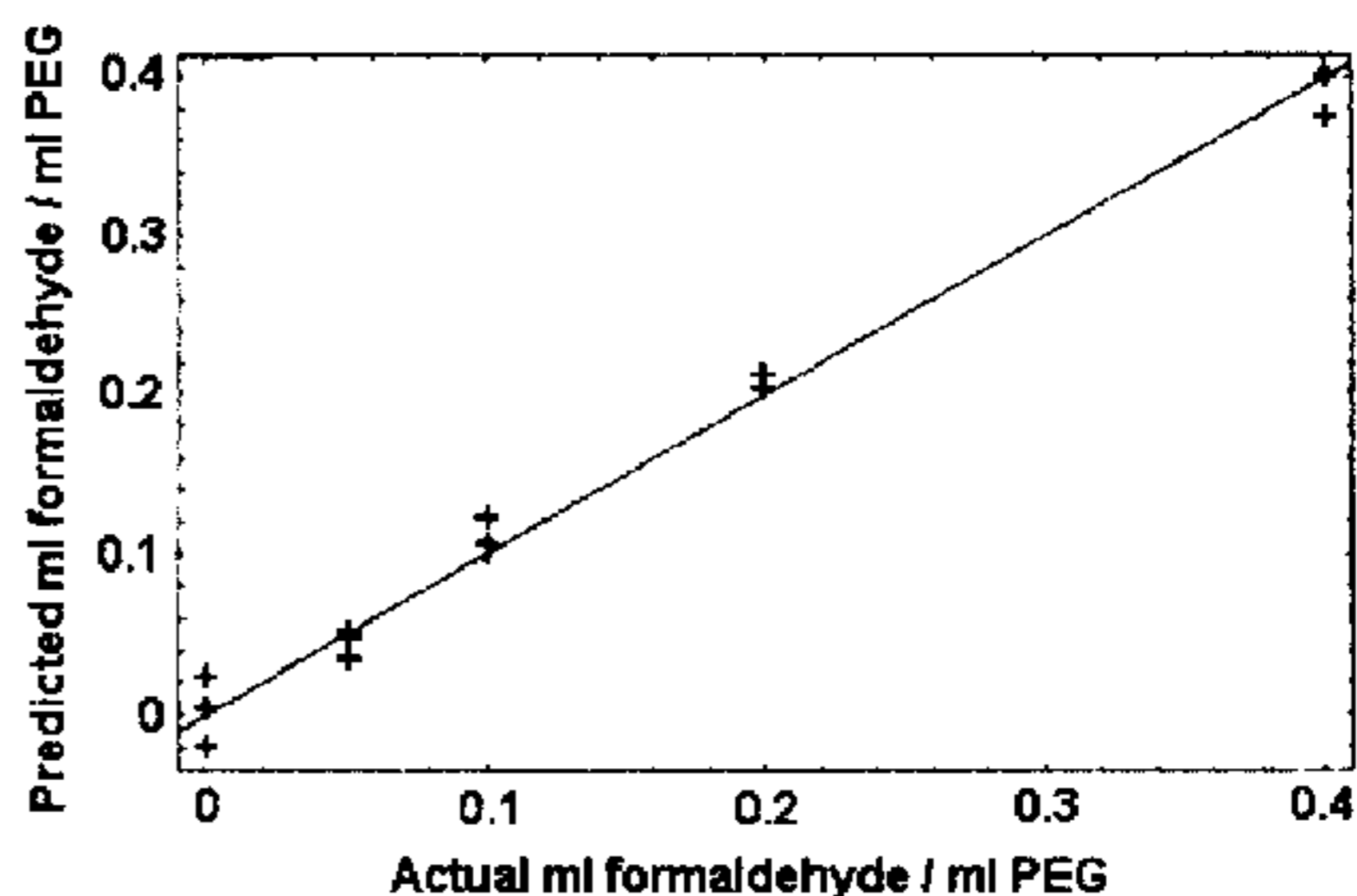


Figure 3 Correlation of actual concentrations of 37% formaldehyde solutions/ml PEG, encapsulated for 48 hours in soft gelatin capsules, to concentrations predicted by NIR spectrophotometry of empty, dry capsules ($r^2 = 0.988$; passed cross-validation, $SEE=0.024$ v/v%, $SEP=0.040$ v/v%; $p<0.05$, f test). Nine of the thirteen generated principal components were utilized in the prediction model. The calibration line, representing $r^2=1$, is shown superimposed on the data points.

capsules, when evacuated. From cross-validation, the standard error of estimate (SEE) and standard error of prediction (SEP) were 0.024 and 0.040%, respectively, with $p < 0.05$ using the f -test. The calibration line, depicting a theoretical correlation coefficient of $r^2=1$, is displayed superimposed over the data points in Figure 3. In effect, the alterations in the NIR spectra of the evacuated gelatin capsules, resulting from formaldehyde migration within and possible reaction with the gelatin shell, were correlated to a recognized concentration of formaldehyde deposited in the fill material of the SEGC. Two possibilities exist that might explain the high correlation between evacuated SEGC NIR spectra and known concentration of formaldehyde doped into the capsules' PEG fill: (1) either all, or, (2) a fixed percentage of the formaldehyde that was originally encapsulated by the SEGCs in each of the experimental groups migrated into and possibly reacted with the gelatin capsule shell. The latter phenomena were demonstrated by concomitant modifications in the NIR spectra of the evacuated SEGCs. If the formaldehyde in the fill of the second and subsequent groups of capsules (formaldehyde solution, 0.10, 0.20, 0.40% v/v in

PEG) had not completely migrated into the gelatin shell of the SEGC, while the formaldehyde in the fill of the first group of capsules (formaldehyde solution, 0.05% v/v in PEG) not only diffused into but also reacted with the gelatin capsule shell, the NIR-predicted value of formaldehyde concentration in the former groups would have been considerably lower than the actual amount of formaldehyde assimilated into the capsule fill. The consequent graph correlating actual concentrations of formaldehyde in PEG, encapsulated for 48 h in soft gelatin capsules, to concentrations predicted by NIR spectrophotometry, would have contained points that diverged downward from the calibration line in Figure 3.

Examination of the vectors of the transformation matrices bridging wavelength and factor hyperspace can reveal the wavelengths (and possible chemical interactions) correlate positively (or negatively) with the NIR spectral changes of the empty SEGCs, initially charged with formaldehyde-tainted PEG. The weightings for the first factor (Figure 4) showed a positive weighting on wavelengths between 1100 and 1400 nm and above 2050 nm. Conversely, the same weightings (Figure 4) displayed negative weights on wavelengths between 1400 nm and 2050 nm. The second weightings graph showed more

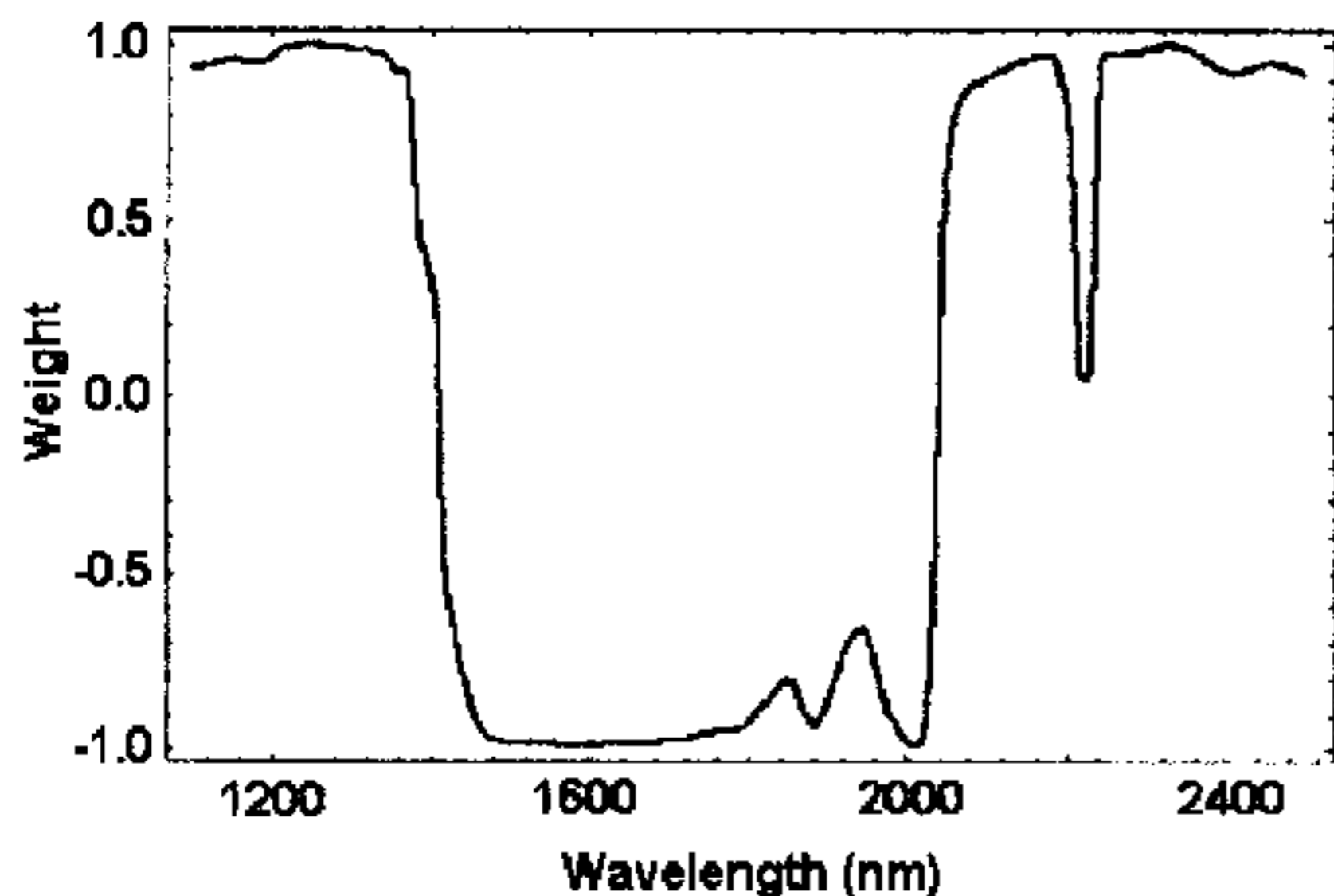


Figure 4 Loadings for the first principal component, showing wavelengths weighted by PCR in the NW spectral changes of empty soft gelatin capsules (filled for 48 h with polyethylene glycol I formaldehyde solution, then emptied and dried).

specific wavelengths of correlation (Figure 5), e.g., large positive weighting on 1385, 1840, and 2200 nm, whereas strong negative weighting was placed on 2040, 1945, and 1420 nm. Of particular significance are the last two wavelengths: 1420 and 1945 nm, which typify the water molecule's NIR absorbance wavelengths. Because water is a product of formaldehyde-induced crosslinking reactions in gelatin (18), the elimination of water from the SEGC gelatin shell with increasing concentrations of formaldehyde in the PEG fill is a plausible explanation for the weightings, with wavelengths representing the NIR signals of water (1945 and 1420 nm; Figure 5). Accordingly, as the concentration

of formaldehyde in the SEGC fill material increased, the NIR spectra of the empty SEGCs, verified by decreases in specific

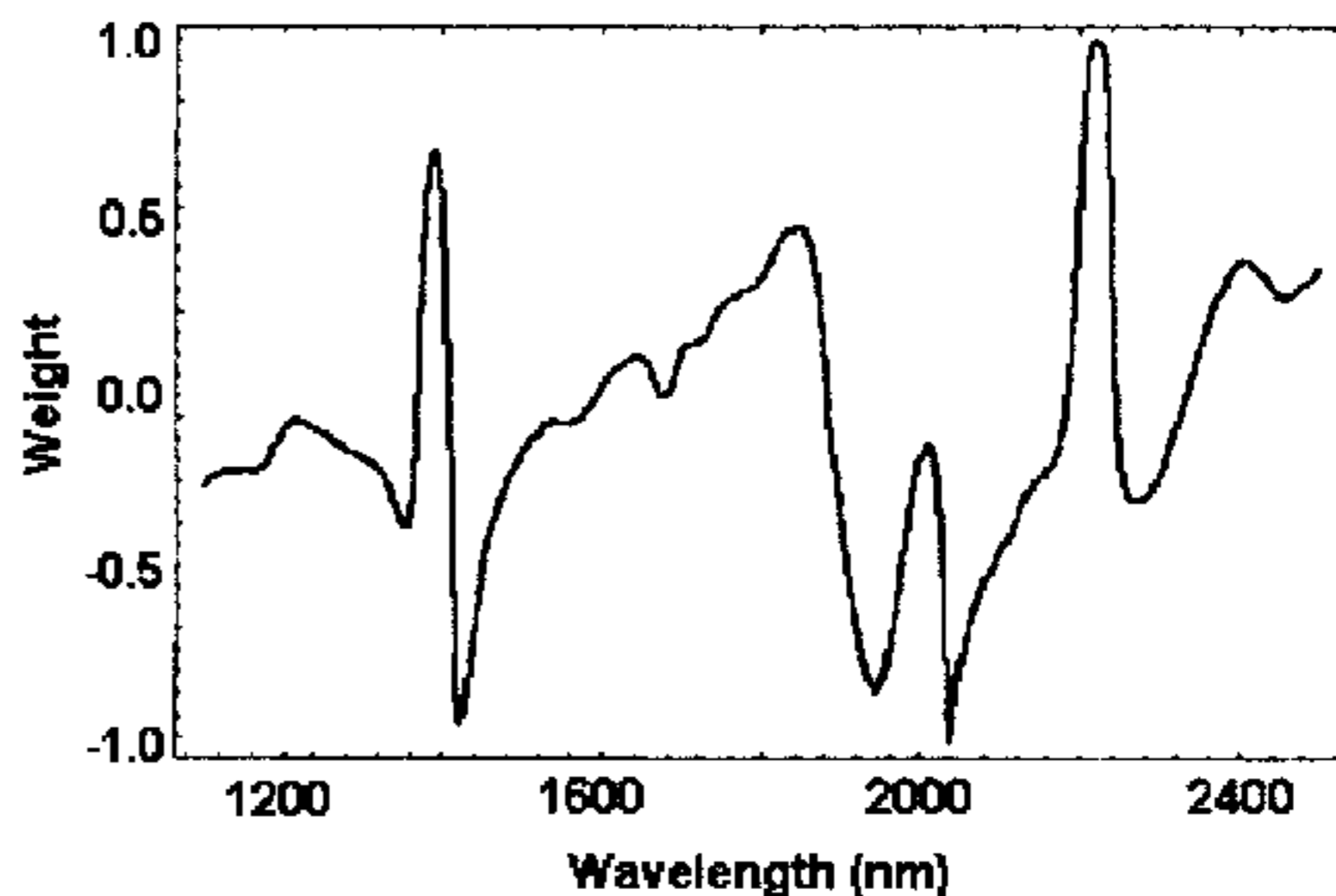


Figure 5 Loadings for the second principal component, showing wavelengths weighted by PCR in the NIR spectral changes of empty soft gelatin capsules (filled for 48 h with polyethylene glycol I formaldehyde solution, then emptied and dried).

spectral regions (1420 and 1945 nm; Figure 5) demonstrated reduction of water by evaporation in the gelatin shell. The weighting of the first factor actually describes a baseline shift in the spectra that arises from a change in water concentration. The absorbance spectrum of water in the NIR region is so strong that a small change in water concentration affects the entire spectral baseline. In the third graph of factor weightings (Figure 6), the appearance of positive wavelength weighting deviations (not due to water absorbance) at 1780 and 2200 nm validates the hypothesis that through the formaldehyde-induced crosslinking of gelatin, chemical bonds are broken and/or formed (1,18). These crosslink bonds give rise to the vibrational harmonic and overtone signal frequencies observed in the NIR region.

Gelatin crosslinking, originated by formaldehyde

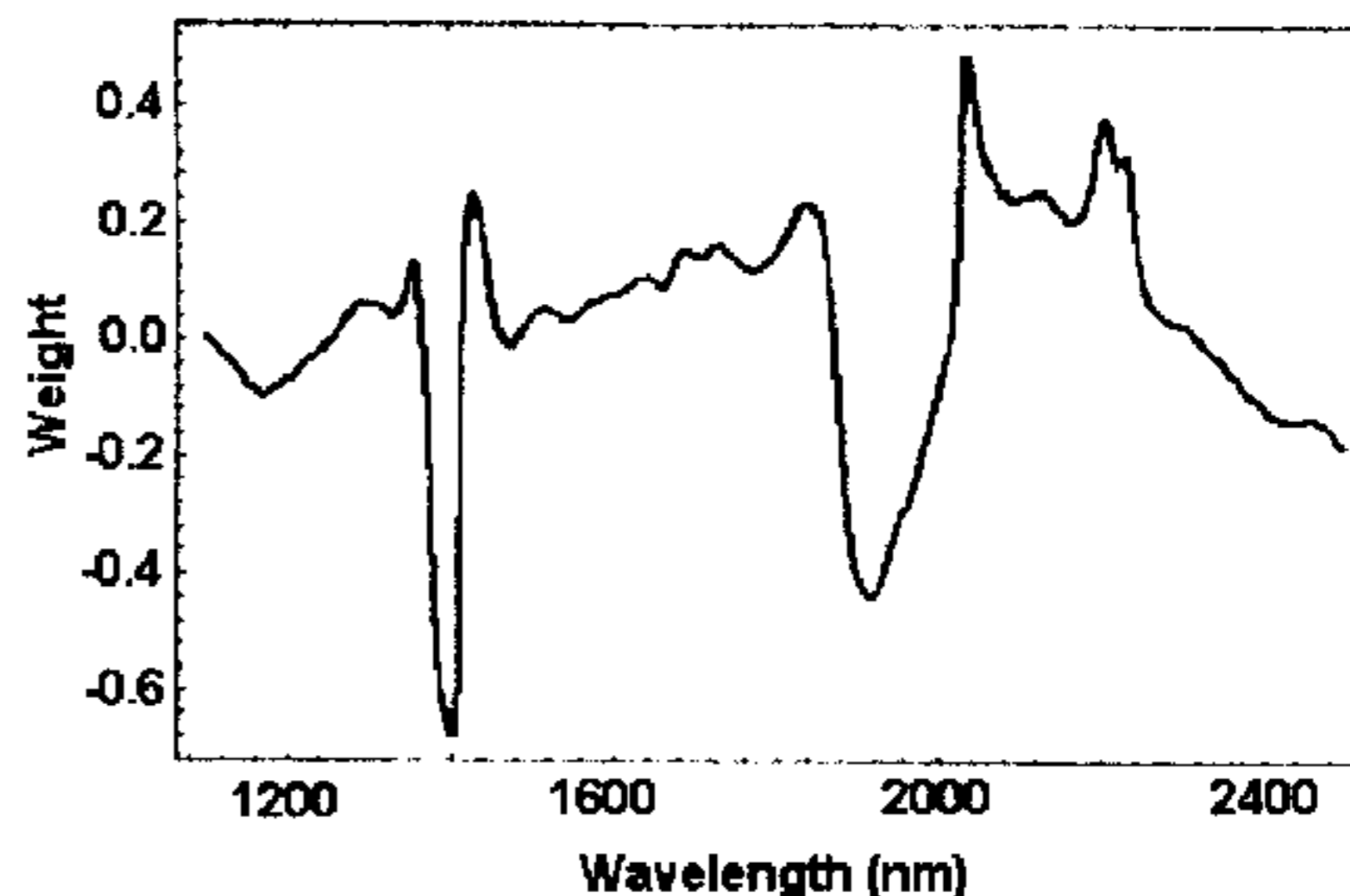


Figure 6 Loadings for the third principal component, showing wavelengths weighted by PCR in the NIR spectral changes of empty soft gelatin capsules (filled for 48 h with polyethylene glycol I formaldehyde solution, then emptied and dried).

introduction into the PEG fill of a SEGC, was observed in the gelatin shell of the SEGC using NIR spectrophotometry. When NIR was coupled to BPS analysis, it was possible to predict the amount of formaldehyde deposited in the original fill material using the NIR spectra of evacuated SEGCs. While not shown in Figure 3, it is probable that substantially higher (and less pharmaceutically relevant) concentrations of formaldehyde in the fill material of SEGCs would have saturated the reactive groups (ϵ -amino and guanidino functionalities of lysine and arginine, respectively) in the gelatin molecule. In this case, NIR would not have been as successful in calculating the amount of formaldehyde initially encapsulated by the SEGC, since much of the formaldehyde would have remained unreacted in the PEG fill and would have been evacuated.

Water content of the SEGC proved to be the largest factor in the variation of capsule spectra representing different amounts of formaldehyde in the PEG fill. NIR spectrophotometry was able to distinguish a depletion of water in the gelatin shell of the same capsule with rising concentration of formaldehyde in the PEG fill of the SEGC. Computer-assisted molecular modeling is in progress to further connect NIR to fundamental molecular motions to further verify proposed gelatin crosslinking mechanisms.

Acknowledgement

The authors thank Drs. Brad Gold and George Digenis for their assistance in beginning gelatin crosslinking experiments, and NSF CHE-9257998 for funding a portion of this research.

REFERENCES

1. G. A. Digenis, T. B. Gold and V. P. Shah. Cross-linking of gelatin capsules and its relevance to their in vitro-in vivo performance. *J. Pharm. Sci.*, 83:915-921(1994).
2. H. Fraenkel-Conrat, M. Cooper and H. S. Olcott. Reaction of formaldehyde with proteins. *J. Am. Chem. Soc.*, 67:950-954 (1945).
3. P. Davis and B. E. Tabor. Kinetic study of the crosslinking of gelatin by formaldehyde and glyoxal. *J. Polym. Sci.*, A1:799-815 (1963).
4. S. K. Taylor F. Davidson and D. W. Ovenall. Carbon-13 nuclear magnetic resonance studies on gelatin crosslinking by formaldehyde. *Photogr. Sci. Eng.*, 22:134-138(1978).
5. K. Albert, B. Peters, E. Bayer, U. Treiber and M. Zwillig. Crosslinking of gelatin with formaldehyde; a ^{13}C NMR study. *Z. Naturforsch.*, 41b:35 1-358 (1986).
6. T. B. Gold, S. L. Smith and G. A. Digenis. Studies on the influence of pH and pancreatin on ^{13}C -formaldehyde-induced gelatin cross-links using nuclear magnetic resonance. *Pharm. Dev. Tech.*, 1:21-26 (1996).
7. G. A. Digenis and T. B. Gold. Chemistry of gelatin cross-linking. *Pharm. Res.*, 11:S146(1994).
8. J. T. Carstensen and C. T. Rhodes. Pellicle formation in gelatin capsules. *Drug Dev. Ind. Pharm.*, 19:2709-2712 (1993).
9. C. B. Bottom, M. Clark and J. T. Carstensen. Dissolution testing of soft shell capsules - acetaminophen and nifedipine. *J. Pharm. Sci.*, 1997 Sep;86(9):1057-61.
10. T. B. Gold, R. G. Buice, Jr., R. A. Lodder and G. A. Digenis. Determination of extent of formaldehyde-induced cross-linking and moisture content of intact hard gelatin capsules by near-infrared spectrophotometry. *Pharm. Res.*, 12:S294 (1995).
11. E. Doelker and A. C. Vial-Bernasconi. Interactions contenant-contenu au sein des capsules gelatineuses et evaluation critique de leurs effets sur la disponibilite des principes actifs, *S. T. P. Pharma.*, 4:298-306 (1988).
12. H. Mohamad, R. Renoux, S. Aiache, and J. M. Aiache. Study on the biopharmaceutical stability of medicines: application to tetracycline hydrochloride capsules I. In vitro study. *S. T. P. Pharma.*, 2:531-535 (1986).
13. T. B. Gold, R. G. Buice, Jr., R. A. Lodder and G. A. Digenis. Determination Of Extent Of Formaldehyde-Induced Crosslinking in Hard Gelatin Capsules By Near-Infrared Spectrophotometry, *Pharm Res* 14(8), 1046-1050, (1997).
14. R. G. Buice, Jr., T. B. Gold, R. A. Lodder and G. A. Digenis. Determination of moisture in intact gelatin capsules by near-infrared spectrophotometry. *Pharm. Res.*, 12:161-163 (1995).
15. Shao, J. Bootstrap Model Selection, *JASA*, 1996, 91(434), 655-665.
16. T. B. Gold, R. G. Buice, Jr., R. A. Lodder and G. A. Digenis. Detection of Formaldehyde-Induced Crosslinking in Soft Elastic Gelatin Capsules Using Near-Infrared Spectrophotometry, *Pharmaceutical Development and Technology*, 3(2), 209-214, 1998.
17. R. A. Lodder, M. Selby, and G. M. Hieftje. Dejection of capsule tampering by near-infrared reflectance analysis. *Anal. Chem.*, 59:1921-1930 (1987).
18. I. D. Robinson. Rate of crosslinking of gelatin in aqueous solution. *J. Appl. Polym. Sci.*, 8:1903-1918 (1964).
18. R. F. Goddi and D. A. Delker. Spectra-structure correlations for the near-infrared region. *Anal. Chem.*, 32:140-141(1960).