Drug Development and Industrial Pharmacy, 26(10), 1045–1057 (2000)

RESEARCH PAPER

Statistical Validation of Reproducibility of HPLC Peptide Mapping for the Identity of an Investigational Drug Compound Based on Principal Component Analysis

Kwan R. Lee,[1,*] Jacob Bongers,[2] Dipti Gulati,[2] and Sudhir Burman[2]

[1]*Statistical Sciences Department and* [2]*Analytical Sciences Department, SmithKline Beecham Pharmaceuticals, King of Prussia, PA 19406*

ABSTRACT

Peptide mapping is a key analytical method for studying the primary structure of proteins. The sensitivity of the peptide map to even the smallest change in the covalent structure of the protein makes it a valuable ''fingerprint'' for identity testing and process monitoring. We recently conducted a full method validation study of an optimized reverse-phase high-performance liquid chromatography (RP-HPLC) tryptic map of a therapeutic anti-CD4 monoclonal antibody. We have used this method routinely for over a year to test production lots for clinical trials and to support bioprocess development. One of the difficulties in the validation of the peptide mapping method is the lack of proper quantitative measures of its reproducibility. A reproducibility study may include method and system precision study, ruggedness study, and robustness study. In this paper, we discuss the use of principal component analysis (PCA) to quantitate peptide maps properly using its projected scores on the reduced dimensions. This approach allowed us not only to summarize the reproducibility study properly, but also to use the method as a diagnostic tool to investigate any troubles in the reproducibility validation process.

Key Words: *Identity testing; Method validation; Monoclonal antibody; Peptide mapping; Principal component analysis; Tryptic mapping.*

* To whom correspondence should be addressed.

INTRODUCTION

Principal component analysis (PCA) in many ways forms the basis for multivariate data analysis. The main function of PCA is to reduce dimensions of the multivariate, multichannel data to a few manageable dimensions. The reduced data serve as an approximation to the original data and allow an analyst to overview (plot) the data in the reduced dimensions and study different cases and variables for their contribution and relationship to overall variability of the data.

Most of the information of the reduced data is contained in the loadings and scores. Loadings are weights given to the original variables and are very useful for locating the important variables. Scores are linear combinations of the original variables and have weights that are given by loadings. Scores are used in the place of the original data and can be interpreted as projected data in the rotated coordinates. The coordinates of the rotation are determined to maximize the capability of capturing the information in the data. Good tutorial articles on PCA are available (1–3).

Peptide mapping is a powerful technique for studying the primary structure of proteins. For recombinant protein pharmaceuticals, peptide mapping is used for the initial ''proof of structure'' characterization, that is, to confirm expression of the desired amino acid sequence and to characterize any posttranslational modifications such as glycosylation, proteolytic processing, acetylation, and so on. Further, peptide mapping is employed for subsequent lot-to-lot identity testing (''fingerprinting'') in support of bioprocess development and clinical trials. Peptide mapping is also the current method of choice for monitoring ''genetic stability.'' Finally, the method is validated to meet stringent federal regulations and is transferred to a manufacturing site for quality control. There are several recent reviews (4–8) on methods and methods validation for peptide mapping as applied to ''well-characterized biopharmaceuticals.''

In recent times, peptide mapping has become much more rapid and convenient as a method (6), and this greatly expands its utility for process monitoring and other higher volume applications. For an analyst to reduce, alkylate, and digest a dozen or more samples in parallel over the course of a working day, nearly impossible by the older process, is now fairly routine. We recently completed a full validation study of an optimized reverse-phase high-performance liquid chromatographic (RP-HPLC) tryptic mapping method for a therapeutic anti-CD4 monoclonal antibody (9).

For the validation of the HPLC peptide mapping, the PCA allow us to measure quantitatively and to summarize the variations in the different maps; hence, it is very useful to study the repeatability and reproducibility of the method in an objective manner (10). In particular, the scores in the first few dimensions can be used in the place of the original peptide maps to study the reproducibility of the method. These are illustrated in this paper on the method precision study and ruggedness (robustness) study of the HPLC peptide mapping method for the identity of a drug compound. As is common in the multivariate data analysis, each variable was centered, that is, the mean of each variable (channel) was subtracted from each instrument reading. In this way, the use of PCA is

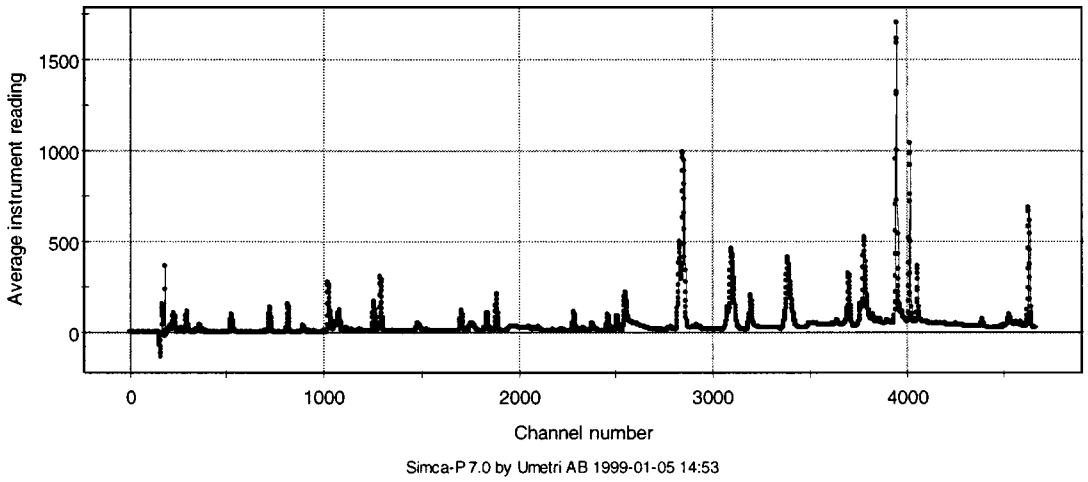

Figure 1. A typical peptide map of 4667 channels.

Table 1

First Two Principal Components and Variations

CompNum	Title, CompName	Lyophile Precision Study, R^2X	R^2X (cum)	Q^2	Q^2 (cum)
1	Comp[1]	0.977	0.977	0.914	0.914
2	Comp[2]	0.020	0.997	0.845	0.986

an attempt to explain the variation of the data from the data centers that is not from the original axis origin. Geometrically centering of data is equivalent to translation of the origin to data centers.

For data analysis, SIMCA® (SIMCA-P for Windows, version 7.0, Umetri AB, Umea, Sweden, 1996) was used for the multivariate analysis and Design Expert® (version 5.0.7, Stat-Ease Corp., Minneapolis, MN, 1997) was used to design and analyze the ruggedness experiment.

METHOD PRECISION AND ITS SYSTEM PRECISION COMPONENT

Method precision (repeatability) was tested by running the entire method for five separate replicate sets of test articles, lyophilized dosage form (Lyo), bulk biological substance (BBS), and reference standard (STD) in parallel (15 samples total). System precision (Rep) was measured separately by running six consecutive RP-HPLC injections from six replicate portions of a single pooled digest of the reference standard. The peptide maps had 4667 instrument readings (absorbance) at as many retention times (Fig. 1). We applied PCA to all 4667 channels to reduce dimensions and to compute the quantitative scores used in the place of the multivariate data.

Precision Studies

Lyophile

Two statistically significant principal components explained most of the variation in the data (Table 1). The R^2 is the fraction of the sum of squares of all the variables explained by the current component. The Q^2 is similar to R^2, but is computed using cross validation. Cross validation is a way of checking the predictive performance of a model on the part of the data not used in the model building. The predicted error sum of squares (PRESS) for the cross validation is the squared differences between observed and predicted values for the data kept out of model fitting. Parts of the data are kept out of the model development, then predicted by the model, and compared

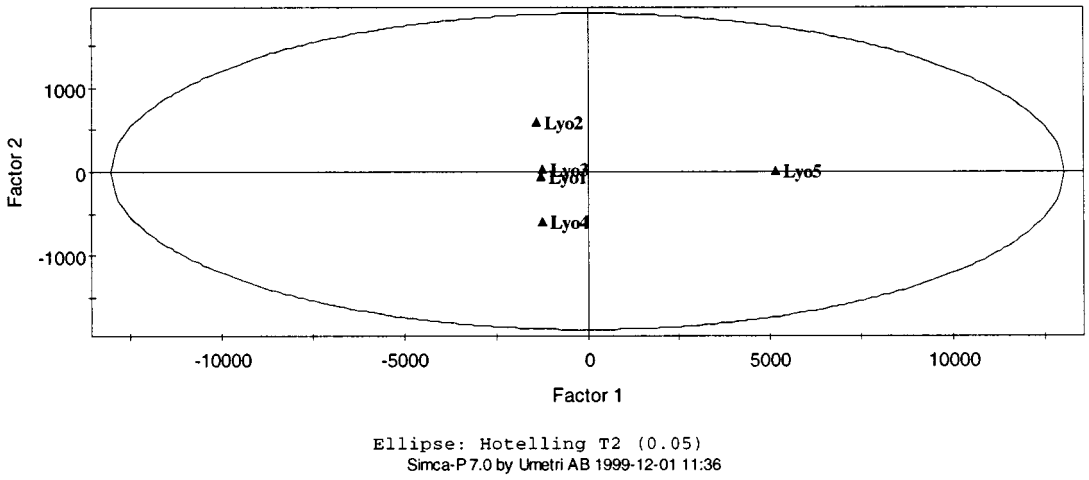

Figure 2. Two-dimensional scores plot of the lyophile precision study.

Table 2

Scores, Tsquare Statistics, and Summary of Precisions

ObsNum	ObsName	Score 1	Score 2	Tsquare	Distance
1	Lyo1	−1284.6	−51.9	0.267	0.516
2	Lyo2	−1406.3	587.5	2.792	1.671
3	Lyo3	−1252.6	33.1	0.243	0.493
4	Lyo4	−1223.1	−585.2	2.699	1.643
5	Lyo5	5166.6	16.5	4.000	2.000
Mean					1.265
Standard deviation					0.708
RSD					55.98%

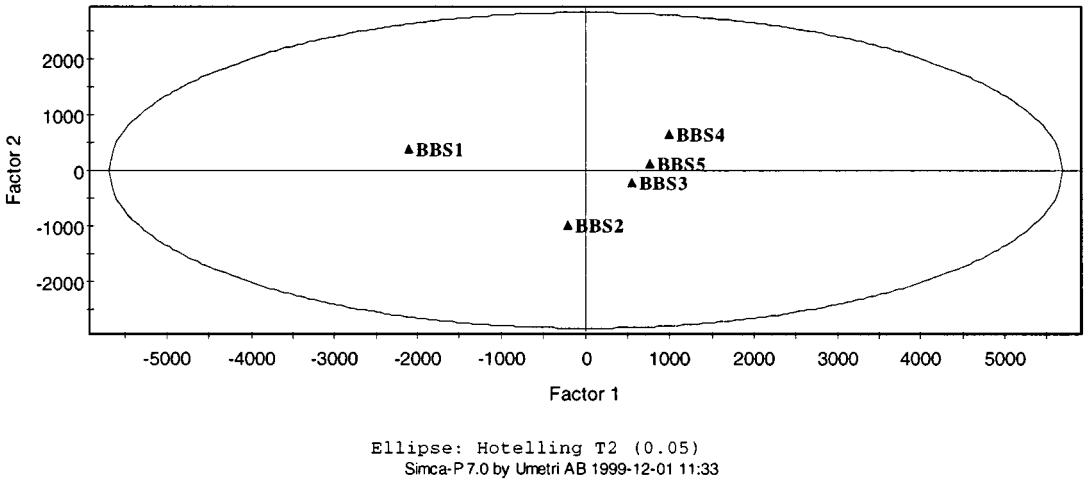

Figure 3. Two-dimensional scores plot of the BBS precision study.

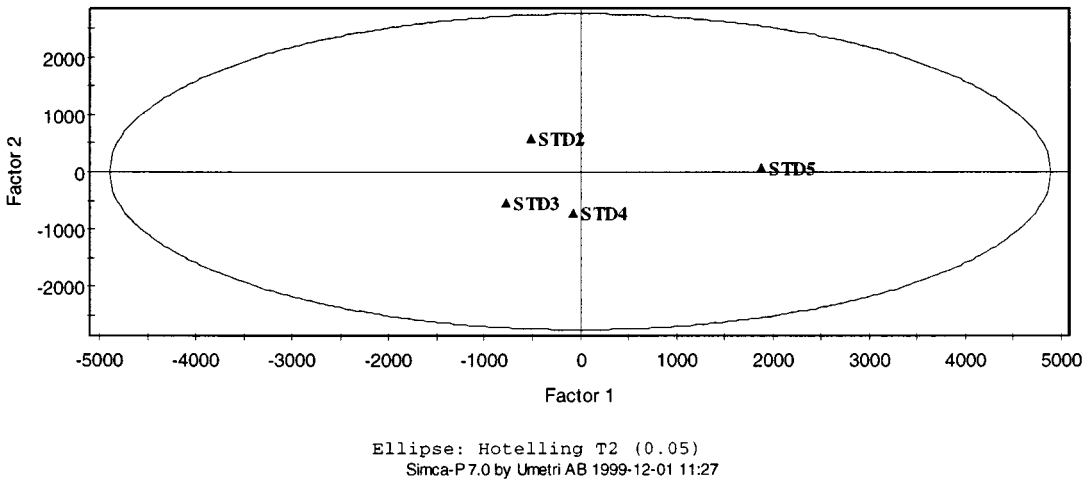

Figure 4. Two-dimensional scores plot of the STD precision study.

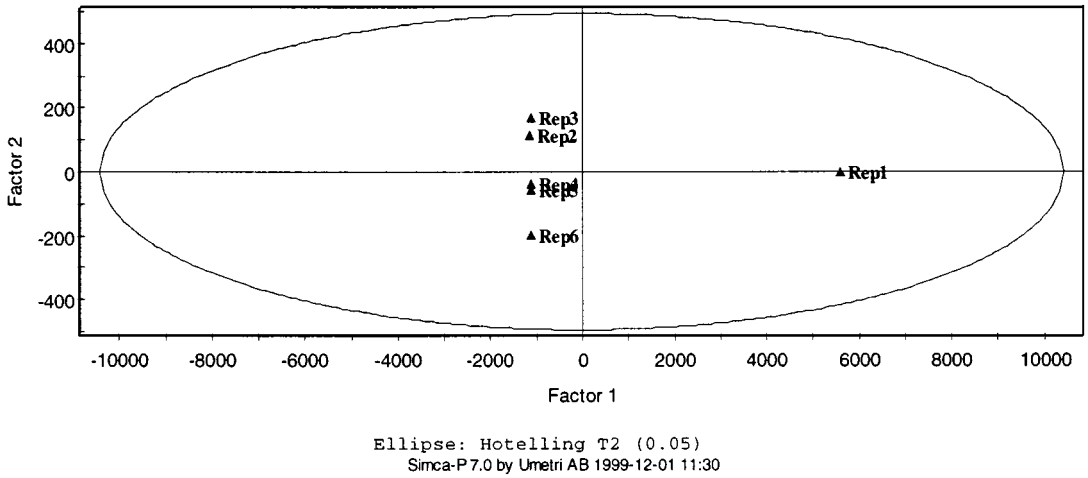

Figure 5. Two-dimensional scores plot of the REP precision study.

with the actual value. This procedure is repeated several times until every data element has been kept out once and only once. The final PRESS then has contributions from all the data. The Q^2 is then computed by 1 − PRESS/SS, where SS is the overall sum of squares. The Q^2 is a more realistic measure of the goodness of approximation than R^2, but Q^2 is not additive like R^2.

The scores plot (Fig. 2) shows the relative position of the five samples from the translated origin (model center), and it can be interpreted as the best two-dimensional projection of the multivariate data. Some data are closer to center than the others; sample Lyo5 is most distant, but is still inside the 95% confidence ellipsoid. If some of the samples are outside the confidence ellipsoid, they can be viewed as possible outliers and may be subject to further investigation.

Tsquare in Table 2 is a combination of the scores in all the projected components (two in this case). It is a

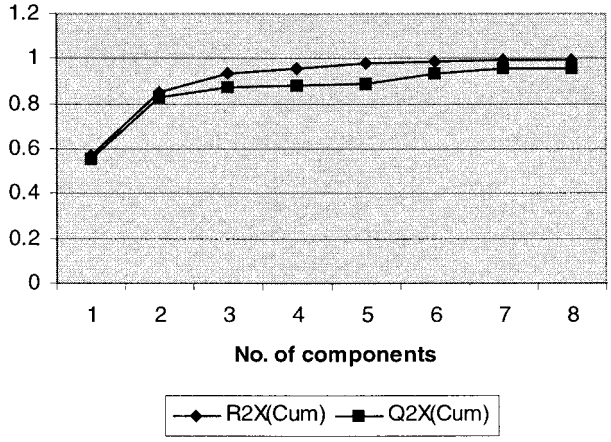

Figure 6. Cumulative variation explained for the merged data.

Table 3

First Eight Principal Components and the Variations Explained

CompNum	CompName	R^2X	R^2X (cum)	Q^2	Q^2 (cum)
1	Comp[1]	0.567	0.567	0.551	0.551
2	Comp[2]	0.281	0.848	0.620	0.829
3	Comp[3]	0.089	0.937	0.270	0.875
4	Comp[4]	0.021	0.959	0.023	0.878
5	Comp[5]	0.016	0.976	0.101	0.890
6	Comp[6]	0.011	0.987	0.391	0.933
7	Comp[7]	0.004	0.992	0.324	0.955
8	Comp[8]	0.002	0.995	0.107	0.960

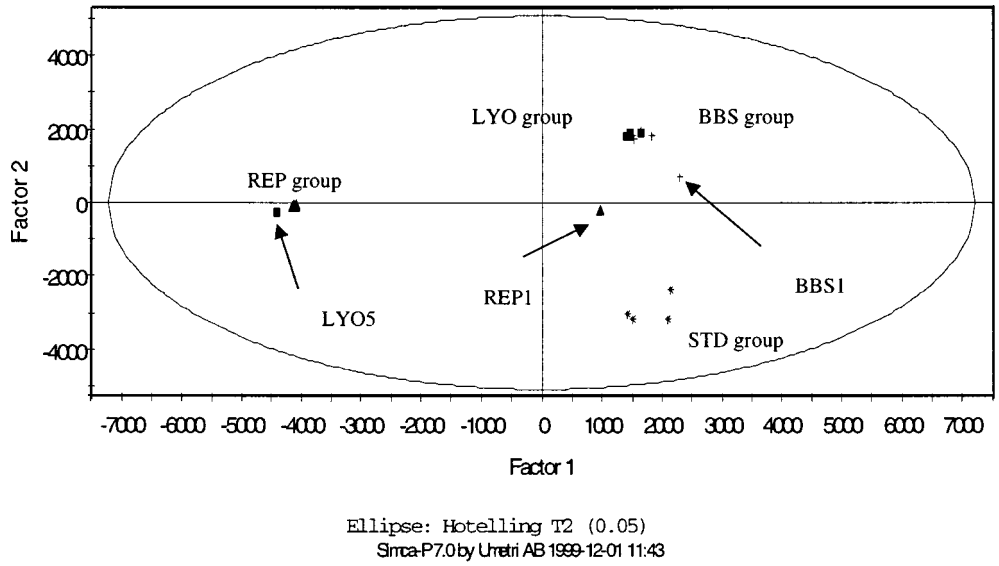

Figure 7. Two-dimensional scores plot for merged data.

scaled measure of how far away an observation is from the center of a model and can be used as a single measure of variability for an observation. *Distance* is the square root of the Tsquare, and it was used to compute the overall summary measure. Also, it should be noted that this distance is measured from the model center, not the distance from the origin, since we have centered the data. Hence, the relative standard deviation (RSD) is not measured against the mean of the original instrument readings; again, it is measured against the mean of the distance from the model center. If different observations have the same distance from the model center, the RSD will be small no matter how large the distance may be. This idea of summarizing the precision studies of HPLC peptide mapping methods using PCA and scores plot is an efficient use of the data, and the summary results on the Tsquare distances can be compared with the other studies later.

Bulk Biological Substance/Reference Standard/System Precision

The same analysis was applied to other precision studies; again, the first two principal components were ade-

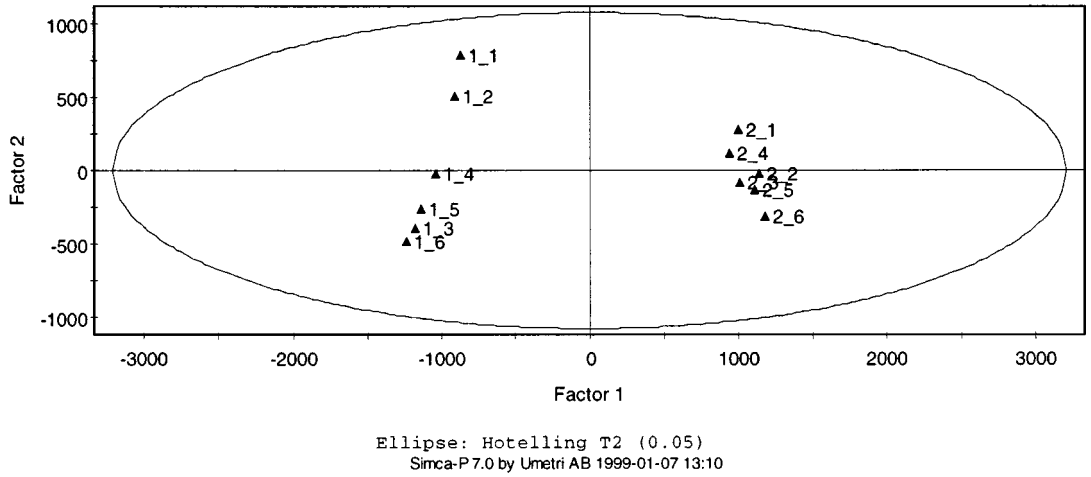

Figure 8. Scores plot of the precision studies on two different instruments.

Table 4

Scores and Distance Measures for the Merged Precision Data

ObsNum	ObsName	Score 1	Score 2	Score 3	Score 4	Tsquare	Distance
1	Lyo1	1457.4	1925.1	557.6	−313.5	2.194	1.481
2	Lyo2	1643.6	1949.2	903.6	−448.0	5.900	2.429
3	Lyo3	1457.7	1846.0	670.4	−338.0	2.377	1.542
4	Lyo4	1395.9	1824.4	256.7	−122.7	2.040	1.428
5	Lyo5	−4423.7	−243.7	600.6	−0.6	18.409	4.291
6	BBS1	2296.8	705.8	−785.2	1939.1	19.923	4.464
7	BBS2	1822.4	1826.0	−644.4	552.5	13.828	3.719
8	BBS3	1638.9	1961.9	105.4	34.6	2.970	1.723
9	BBS4	1536.0	1755.8	630.1	−396.9	3.846	1.961
10	BBS5	1504.0	1837.8	314.6	−148.8	1.888	1.374
11	STD1	2094.6	−3161.9	210.2	165.2	8.295	2.880
12	STD2	2094.6	−3161.9	210.2	165.2	8.295	2.880
13	STD3	1413.5	−3017.6	−6.3	12.8	6.343	2.519
14	STD4	1512.8	−3144.2	658.5	−309.7	16.651	4.081
15	STD5	2134.0	−2352.0	839.7	−356.7	19.448	4.410
16	Rep1	976.4	−182.5	−4297.4	−805.3	19.971	4.469
17	Rep2	−4149.3	−80.7	−45.6	78.0	2.979	1.726
18	Rep3	−4127.5	−69.7	−64.6	68.6	3.189	1.786
19	Rep4	−4110.0	−87.6	−32.7	59.6	2.997	1.731
20	Rep5	−4079.7	−58.7	−38.2	79.0	3.363	1.834
21	Rep6	−4088.5	−71.7	−43.4	85.6	3.094	1.759
Mean							2.595
Standard deviation							1.154
RSD							44.48%

quate to describe the most variation in each of the studies. The R^2 were 0.966, 0.897, and 0.998 for BBS, STD, and REP, respectively. Corresponding Q^2 were 0.856, 0.710, and 0.935, respectively.

The two-dimensional projections of the data are shown (Figs. 3–5). The three samples BBS1, Lyo5, and Rep1 were distinctly different from the others and exerted most of the variation. We have subsequently found that they had shifted retention times. The overall summary measures (RSDs) were given as 46.38%, 23.04%, and 35%, respectively.

Data Analysis of All Four Precision Studies Together

We dealt with each of the four precision studies separately in the analysis above. By treating them as one data set, we can compare them better, both graphically and analytically, for the study of ''between'' precision studies in addition to ''within'' precision studies.

The first eight principal components were statistically significant for explaining variation in the data (Table 3), but it appears that the first three to four components would be adequate for explaining most of the variability (Fig. 6).

As indicated above, we found that the samples BBS1 and Rep1 had shifted retention times and were separated from the other groups (Fig. 7). Also, the four different

Table 5

Distance Measures by Experiment

Trial	Lyo	BBS	STD	Rep
1	1.48		2.88	
2	2.43	3.72	2.88	1.73
3	1.54	1.72	2.52	1.79
4	1.43	1.96	4.08	1.73
5	4.29	1.37	4.41	1.83
6				1.76
Average	2.23	2.19	3.35	1.77
Standard deviation	1.22	1.04	0.84	0.04
RSD (%)	54.65	47.59	24.90	2.51

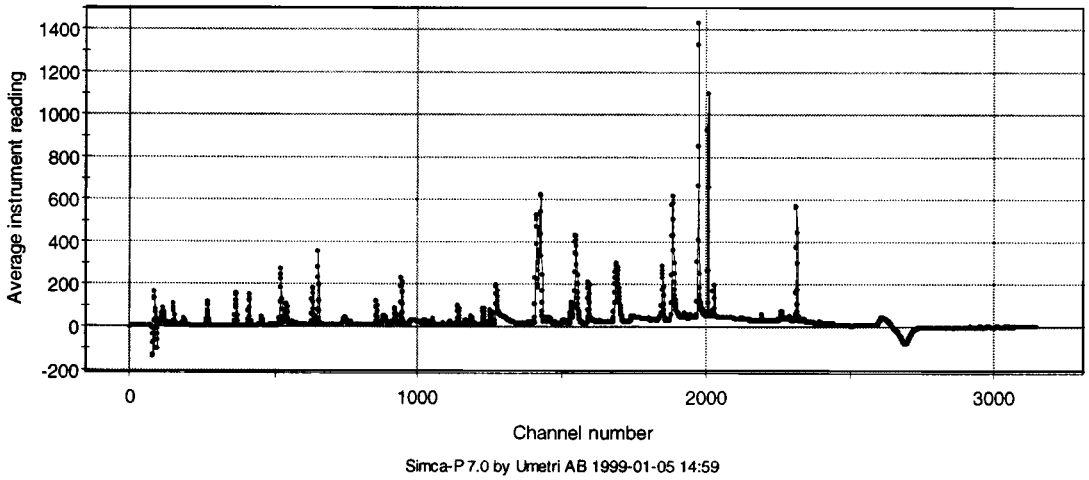

Figure 9. A typical peptide map used in the ruggedness study.

Table 6

Components of Variance for the Merged Precision Study ANOVA After Deleting Samples Lyo5, BBS1, Rep1

	df	MS	Varcomp	Variation (%)
Between	3	3.8638	0.8585	98.14
Within	14	0.0163	0.0163	1.86
Total	17		0.8748	100.00

experiments formed clusters. This experiment-to-experiment variability reflects the difference in samples, which is mostly due to variability in instrument setup. We currently are investigating the possibility of using an alignment algorithm (10) to get rid of the chromatographic variations as a prerequisite to validation studies. An effective alignment algorithm should get rid of the chromatographic variations so that we can study the variations due to sample compositions more accurately.

The following scores plot (Fig. 8) shows very clearly the effects of chromatographic variations from instruments separately from the sample variations. Although the samples in the figure came from a different study, the same pooled digests, mobile phases, columns, and so on were used across the experiment. The only variable was the different instrument used for each trial. It appears that the first factor explains the instrument-to-instrument variability, while the second factor accounts for replicate-to-replicate variability. We expect a good alignment algorithm will effectively remove the instrument-to-instrument variability so that the two clusters in the figure would merge.

The first four component scores and corresponding

Table 7

Factors of the Ruggedness Experiment

Name	Unit	Type	Low	High
A: DDT	μmol	Factor	22.50	27.50
B: Sodium iodoacetate	μmol	Factor	54.00	66.00
C: Reduction time	min	Factor	20.00	40.00
D: Alkylation time	min	Factor	30.00	50.00
E: Substrate/enzyme (wt/wt)		Factor	90.00	110.00
F: Digestion temperature	°C	Factor	35.00	39.00
G: Digestion time	min	Factor	90.00	150.00

Table 8

Design Specifications for the Ruggedness Experiment

Run	A	B	C	D	E	F	G
1	22.50	66.00	40.00	30.00	90.00	39.00	90.00
2	25.00	60.00	30.00	40.00	100.00	37.00	120.00
3	22.50	54.00	40.00	50.00	90.00	35.00	150.00
4	27.50	66.00	20.00	50.00	90.00	35.00	90.00
5	22.50	54.00	20.00	50.00	110.00	39.00	90.00
6	22.50	66.00	20.00	30.00	110.00	35.00	150.00
7	27.50	54.00	40.00	30.00	110.00	35.00	90.00
8	27.50	66.00	40.00	50.00	110.00	39.00	150.00
9	27.50	54.00	20.00	30.00	90.00	39.00	150.00

Table 9

First Four Components of the Ruggedness Experiment

A	R^2X	R^2X (cum)	Eigen Value	Q^2	Q^2 (cum)
1	0.626	0.626	11.270	0.579	0.579
2	0.244	0.870	4.390	0.591	0.828
3	0.038	0.908	0.692	0.164	0.856
4	0.026	0.935	0.473	0.119	0.873

distance measures were computed (Table 4) and summary measures given. After deleting the three readings (to see how much reduction of error we could make) we identified as the results of possible shift of retention times, the within-experiment variabilities were summarized (Table 5). Using analysis of variance to separate the sources of variation, we obtained the results in Table 6. Again, three ''outliers'' were removed from the data before analysis. The percentage of variation explained by between-experiment variation is 98%, and the variation due to within-experiment variation is 2%. After proper alignment of chromatograms, we expect to eliminate most of the variation in the data.

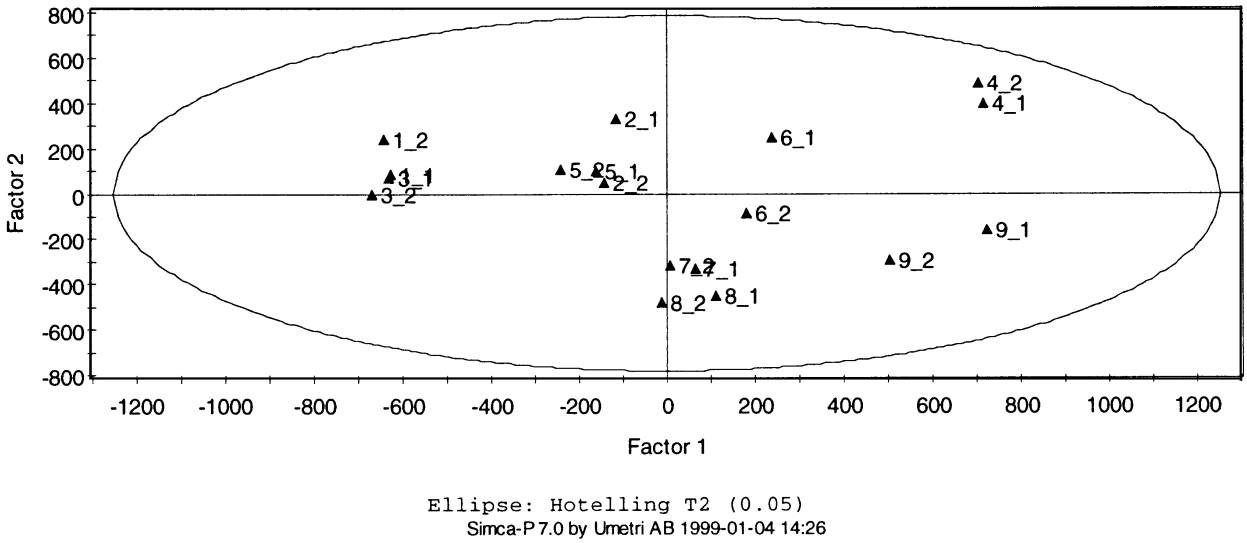

Figure 10. Two-dimensional scores plot of the ruggedness study.

Table 10

First Four Scores for Ruggedness Study

ObsNum	ObsName	Score 1	Score 2	Score 3	Score 4
1	1_1	−628.61	88.84	64.65	149.16
2	1_2	−643.86	240.59	30.76	131.82
3	2_1	−113.79	330.89	−69.04	−67.71
4	2_2	−143.60	47.30	−72.46	−26.86
5	3_1	−632.01	68.71	−24.67	−48.39
6	3_2	−669.31	−0.97	12.50	−31.18
7	4_1	716.14	395.66	132.76	17.67
8	4_2	705.08	486.91	127.82	35.83
9	5_1	−159.69	102.60	−11.51	−32.88
10	5_2	−243.20	108.21	−89.20	29.38
11	6_1	236.65	250.98	−166.62	−142.89
12	6_2	181.76	−83.89	−63.01	−159.74
13	7_1	64.25	−328.76	172.59	−86.33
14	7_2	9.28	−319.54	231.32	13.84
15	8_1	110.28	−453.24	52.04	−83.06
16	8_2	−13.01	−481.33	−68.20	28.84
17	9_1	721.22	−158.64	−56.00	93.63
18	9_2	502.44	−294.32	−203.74	178.89

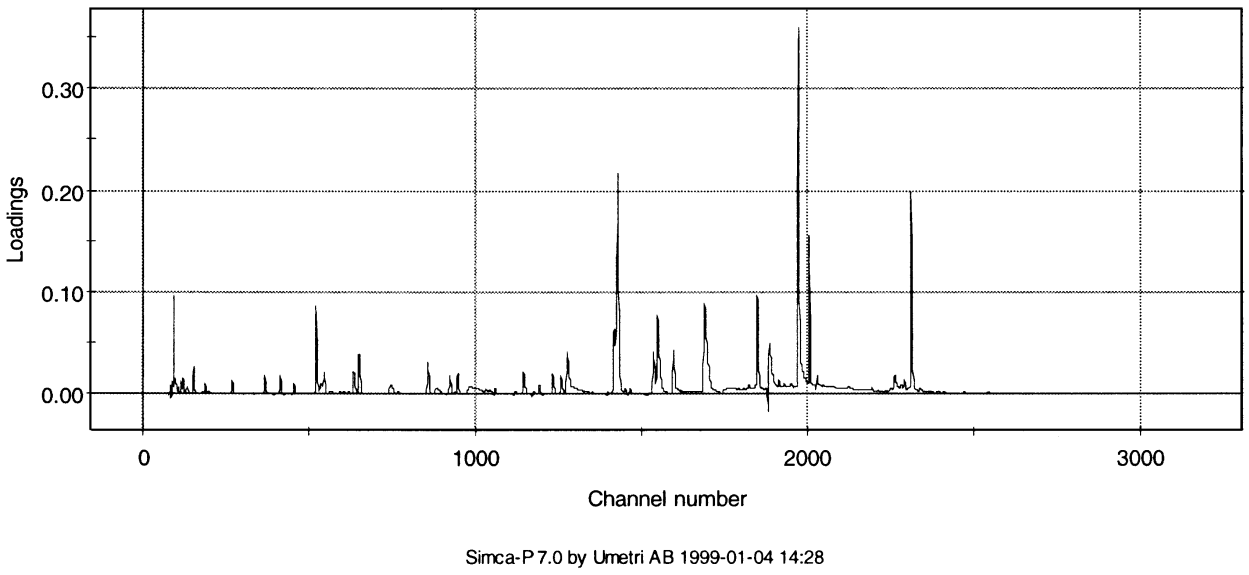

Figure 11. The first loadings plot of the ruggedness study.

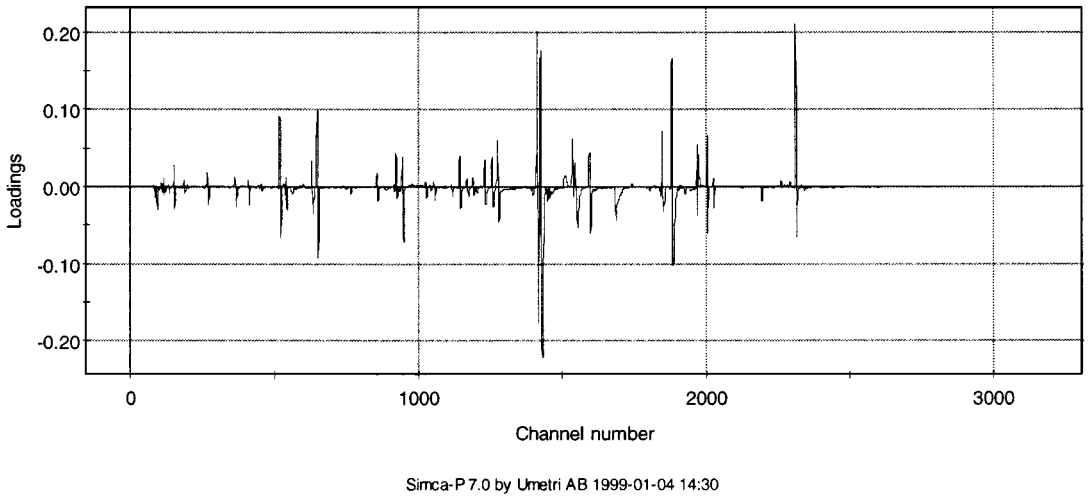

Figure 12. The second loadings plot of the ruggedness study.

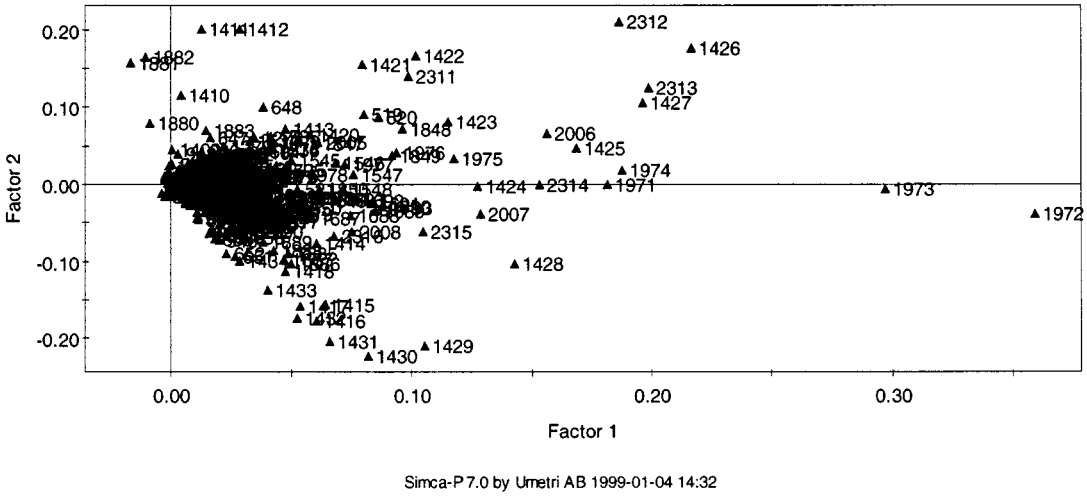

Figure 13. The first two loadings of the ruggedness study.

Table 11

First Two Scores Averaged Over the Two Replicate Injections

	Score 1			Score 2		
Run	Injection 1	Injection 2	Average	Injection 1	Injection 2	Average
1	−628.61	−643.86	−636.24	88.84	240.59	164.72
2	−113.79	−143.60	−128.70	330.89	47.30	189.10
3	−632.01	−669.31	−650.66	68.71	−0.97	33.87
4	716.14	705.08	710.61	395.66	486.91	441.29
5	−159.69	−243.20	−201.45	102.60	108.21	105.41
6	236.65	181.76	209.21	250.98	−83.89	83.55
7	64.25	9.28	36.77	−328.76	−319.54	−324.15
8	110.28	−13.01	48.64	−453.24	−481.33	−467.29
9	721.22	502.44	611.83	−158.64	−294.32	−226.48

RUGGEDNESS AND ROBUSTNESS STUDY

The peptide maps for ruggedness study had a slightly smaller number of time points measured than the precision study; a typical map is shown in Fig. 9. The ruggedness of the study was evaluated for day-to-day, analyst-to-analyst, laboratory-to-laboratory reproducibility. The lot-to-lot consistency of the column and the trypsin were also studied, and suitable alternative vendors for these two critical materials were identified. An interlaboratory study was also conducted with the quality control laboratory at the manufacturing site. We validated the robustness of the method when challenged with the perturbations of selected key operating parameters from the reduction/carboxymethylation/digestion portion (Table 7).

Designed Experiment of Ruggedness Study

The ruggedness experiment was designed using the seven factors listed in Table 7. The fractional factorial was 1/16 of the full two-level, seven-factor factorials with 1 center point, which resulted in 9 runs. Each of the 9 runs was replicated. Table 8 shows the design specifications of the 9 runs in a random order.

Principal Component Analysis of Ruggedness Study

The first four components were statistically significant (Table 9), but the first two components explained more than 80% of the variation. The scores plot (Fig. 10) and individual scores are given below (Table 10). Replicates were close together, and most of the variations came from the sample-to-sample variation.

The first loadings plot is similar to the peptide map in general and seems to be averaging the information in the channels (Fig. 11); the second loadings plot shows the deviation from the averages (Fig. 12). The loadings scatter plot of the first two factors (Fig. 13) clearly shows the important channels of the projection; channels 1972–1974 are more important than the others and basically provide the largest peak in the map. Since the data were not scaled, this is largely expected.

Test of the Factor Effects for Significance in the Ruggedness Study

We have tested the factor effects against the pure errors computed from the replicates and found that most of the factors were statistically significant. In other words, the changes due to factors were larger than underlying noise. However, it was later found that the replicates were not generated properly in the experiment. Instead of using a random order of experiments, the replicates were merely two duplicate injections. As a result, the replicate-to-replicate variation was very small, and all the factors were seen to be statistically significant. In other words, the noise was not estimated properly since the duplicate injections did not reflect all the system noise possible in the experiment. A statistically more reliable procedure for analyzing this experiment would be to use the averages of the duplicate injections and estimate the system noise from the insignificant factors. The averages of two injections were computed for each of the nine runs (Table 11), and these nine runs were used as if they were obtained from a single experiment.

The analysis of the first score showed that the two factors DTT-amount (A) and reduction (C) significantly af-

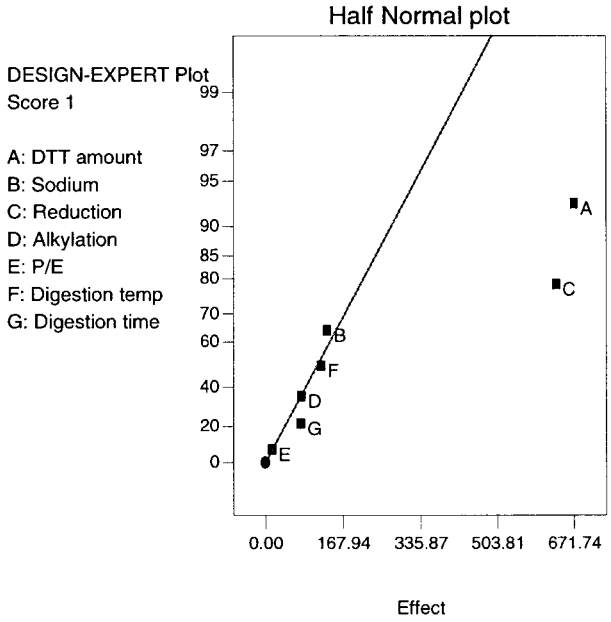

Figure 14. Half-normal plot of the factors for the first scores.

Table 12

T Scores and p Values for Significant Factors for the First Score

Factor	Estimate	t(Coeff = 0)	Probability > $\|t\|$
A: DTT amount	335.87	7.09	.0009
C: Reduction	−316.46	−6.68	.0011

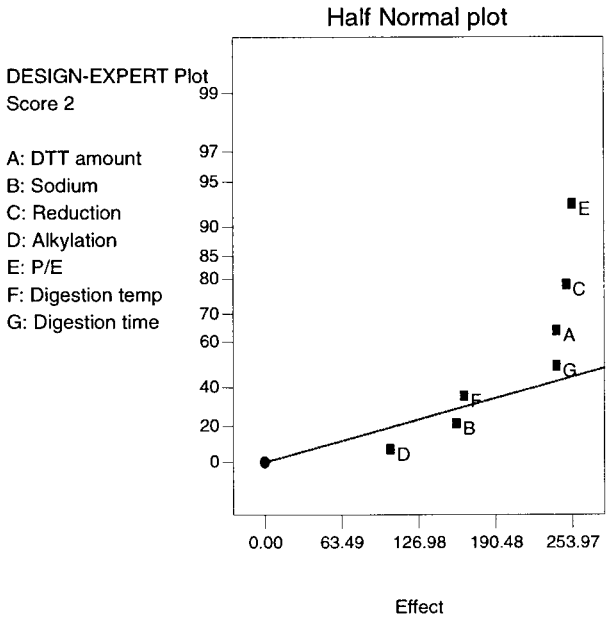

Figure 15. Half-normal plot of factors for the second score.

fected the score. The half-normal plot of the factors for the first score showed clearly the significance of the two factors (Fig. 14). The *p* values of the significant factors from the *t* test are also given (Table 12). None of the factors was significant for the second scores, however (Fig. 15). It appears that we need to control the two factors more tightly along the middle specification of factors.

CONCLUSION

We have shown that principal component analysis is useful for analyzing the unintegrated chromotographic raw data. The scores plot and summary measures such as scores and Tsquares are valuable for method validation, especially for precision study and ruggedness study. The reduced data serve as an approximation to the original data and allow an analyst to overview (plot) the data in the reduced dimensions and to study different cases and variables for their contribution and relationship to overall variability of the data. Excellent diagnostic information may be obtained that can be used directly to improve the precision and ruggedness of the peptide mapping method.

REFERENCES

1. J. E. Jackson, *A Users Guide to Principal Components*, Wiley, New York, 1991.
2. S. Wold, Chemometrics Intell. Lab. Syst., 2, 37–52 (1987).
3. S. Wold, Technometrics, 20, 397 (1978).
4. R. L. Garnick, N. J. Solli, and P. A. Papa, Anal. Chem., 60, 2546–2557 (1988).
5. M. M. Federici, Biologicals, 22, 151–159 (1994).
6. S. A. Carr, M. E. Hemling, M. F. Bean, and G. D. Roberts, Anal. Chem., 63, 2802–2824 (1991).
7. J. J. Dougherty, Jr., L. M. Snyder, R. L. Sinclair, and R. H. Robins, Anal. Biochem., 190, 7–20 (1990).
8. E. R. Hoff and R. C. Chloupek, Meth. Enzymol., 51–68 (1996).
9. J. Bongers, J. J. Cummings, M. B. Ebert, M. M. Federici, L. Gledhill, D. Gulati, G. M. Hilliard, B. H. Jones, K. R. Lee, J. Mozdz̄anowski, M. Naimoli, and S. Burman, J. Pharm. Biomed. Anal., 21, 1099–1128 (2000).
10. G. Malmquist and R. Danielsson, J. Chromatogr., 687, 71–88 (1994).