Use of Cumulative Distribution Functions to Characterize Mass Spectra of Intact Proteins

Paul S. Blank, Christin M. Sjomeling, Peter S. Backlund, and Alfred L. Yergey

Laboratory of Cellular and Molecular Biophysics, National Institute of Child Health and Human Development, Bethesda, Maryland, USA

The MH^+ ions of matrix assisted laser desorption ionization time-of-flight (MALDI-TOF) spectra for a series of closely related but otherwise indistinguishable proteins were analyzed for singularity using a distribution free statistic, the Kolmogorov-Smirnov non-parametric statistic, K–S. The approach allows spectra which might otherwise be taken as identical, to be distinguished. Such analysis of the spectra may lead to a greater understanding of the chemistry of the proteins under investigation. The analysis requires only standard instrumentation. A standard data analysis protocol was developed and applied to generate a normalized cumulative distribution function (NCDF) for each spectrum. Differences in the NCDF for two different spectra were calculated and the maximum difference, Δ_{max} compared to critical values of K–S. Values of Δ_{max} exceeding the critical value of K–S are taken as the basis for rejecting the statistical null-hypothesis and assigning statistical significance to the differences in the two spectra. We have shown that this approach allows spectra of 1:1 mixtures of closely related recombinant proteins to be distinguished from either protein alone, and that mixtures of a 45 kDa protein and a labeled version of that protein can be distinguished from the pure material and from one another at the level of about 25%. In addition, we are able to use this approach to characterize the extent to which a synthetic glyococonjugation reaction has proceeded under circumstances of differing reaction times. (J Am Soc Mass Spectrom 2002, 13, 40–46) © 2002 American Society for Mass Spectrometry

Successful solutions to numerous problems in biology, biochemistry, and biotechnology depend on the ability to produce "pure" proteins or to recognize the degree to which these molecules might be modified. Widely used methods for assessing purity such as polyacrylamide gel electrophoresis (PAGE), size exclusion chromatography (SEC), or high performance liquid chromatography (HPLC) are relatively nonspecific and insensitive to small differences in molecular weight [1]. Mass spectrometric based peptide mapping methods such as MALDI-TOF [2, 3] or liquid chromatography tandem mass spectrometry (LC/MS-MS) [4–6], while very effective in identifying sites and types of modifications, are limited in their ability to assess protein purity or the extent of protein modifications. However, mass spectrometric analysis of the spectra of intact proteins offers the possibility of solving these problems.

Two factors limit the possibility of MS being used for analysis of large intact molecules. The first is most obviously the resolving power of the mass spectrometer. In general, the better the resolution that can be achieved, the more effectively the problems of interest can be addressed. In a practical sense however, the performance of current widely available instruments limits the determination of protein purity or the presence of modifications to molecules with molecular weights $\leq$20 kDa. When considering the mass spectra of even larger molecules, diminished resolution and the increased breadth of the isotope cluster further complicate the analysis. For example, a mathematical simulation of the isotope clusters of two proteins of molecular weight 16 and 66 kDa at a fixed resolving power [7] shows that a 66 kDa molecule has an isotope cluster more than four times broader than the 16 kDa molecule. The peak widths are 15 versus 70 Da at resolving power of 1000. While resolving powers of about 1000 can be readily attained at 20 kDa with existing high performance MALDI-TOF instruments, at 66 kDa the resolving power drops to a few hundred. Therefore, although direct observation of mixtures that differ only slightly in mass is possible in the region of 20 kDa, the presence of mixtures containing the same small differences in mass would not likely be detected for 66 kDa molecules. Similar arguments hold for electrospray ionization (ESI)

Published online November 19, 2001

Address reprint requests to Dr. A. L. Yergey, Building 10, Room 9D52, MSC 1580, Laboratory of Cellular and Molecular Biophysics, NIH, Bethesda, MD 20892, USA. E-mail: aly@helix.nih.gov

1044-0305/02/$20.00
PII S1044-0305(01)00338-5

Received July 9, 2001
Revised September 26, 2001
Accepted September 26, 2001

spectra of large molecules because of the high charge states that must exist in order to obtain spectra within the mass range of commonly available analyzers.

As an alternative approach to solving these problems, we have developed a nonparametric statistical analysis of the parent ion region of the mass spectra of intact proteins. The analysis shows great promise in assessing protein heterogeneity and in detecting otherwise indistinguishable differences in compounds of very similar molecular weights. A principal advantage to our approach is that no additional instrumentation is required. We have applied this technique to mass spectra of proteins generated using MALDI-TOF. In principle, however, the method could be applied to spectra obtained using any mass spectrometer employing any ionization scheme.

Experimental

Instrumentation

MALDI. Mass spectra were collected on a PE-Biosystems (Framingham, MA) Voyager DE-STR instrument operated in linear mode with delayed extraction. General instrument operating parameters were 25 kV accelerating voltage, Va, delay time: 400 ns, grid voltage: 90% of Va, and guide wire voltage: 0.3% of Va. Grid voltage and delay time were varied according to analyte molecular weight in order to optimize resolution of the MH^+ ion. Mass analysis operating parameters were kept constant for each compound type. Specta were acquired using a Tektronix (Beaverton, OR) Model 540B 2 GS/s digital oscilloscope and were transmitted to GRAMS (Galactic Industries, Salem, NH) spectral analysis package following acquisition. Nitrogen laser intensities were varied to optimize signal intensities without sacrificing resolution, and typically varied between 2600 and 2900 arbitrary units of intensity. Typically the default mass calibration file was used for mass assignment.

Samples

The samples used in this study were analyzed in the course of the normal operation of the NICHD Mass Spectrometry Facility as part of its ongoing role within the Institute. All samples were presented simply for determination of their molecular weights. The samples were delivered either in solution or in lyophylized form. The recombinant α- and β-tubulins were supplied by D. Sackett, LIMB, NICHD, the synthetic glycoconjugates by V. Pozsgay, LDMI, NICHD, and the fluorescent tagged proteins by S. Yefimov, SMA, NICHD.

Sample Preparation

All samples were prepared using a three layer deposition on the MALDI plate of matrix sample matrix. This approach allowed the removal of salts by subsequent washing between application of sample and the second application of matrix. The first matrix application consisted of a saturated solution of sinnapinic acid (3,5-Dimethoxy-4-hydroxy-cinnamic acid, Aldrich) in acetone. The acetone led to rapid formation of small matrix crystals. The source of the top matrix layer was a saturated solution of sinnapinic acid in equal volumes of acetonitrile and 0.1% trifluoracetic acid, TFA (Fluka, Milwaukee, WI). Protein samples were dissolved in either water or 0.1% TFA at concentrations $\geq$2 pmole/μL. For each layer of the sample, a volume of approximately 0.5 μL was applied and allowed to dry in air. Samples that did not yield adequate MALDI spectra, presumably because of high salt levels, were washed by applying and removing after 15 s, 10μL of 0.1% TFA to the dried sample spot. Matrix was not reapplied after washing.

Statistical Analysis of Spectra

A generalized analysis of peak shape was used to distinguish between the MH^+ ions of closely related compounds. A nonparametric, distribution free statistic is used. Stated simply, a nonparametric statistic makes no assumptions about the underlying shape of the distributions being analyzed, but compares them in a rigorous and unbiased fashion. As in other statistical determinations, the object of the comparison is to test the validity of the null hypothesis, H_0. That is, within the limits of the statistical test, does some parameter equal or exceed a critial value? If so, then the null hypothesis is rejected and the test being conducted is assigned a statistical significance based on the limits of the critical parameter. We use the two-sided, distribution-free Kolmogorov-Smirnov K–S statistic for large n [8, 9].

In this work, spectra are smoothed using a 19 point, second-order moving average (Savitzy-Golay) algorithm that is enabled in the GRAMS package. The molecular ion MH^+ region of the full mass spectrum is chosen for subsequent analysis. This region is made as narrow as possible without cutting off the tails of the actual peak. Truncation of the mass spectra is accomplished in two steps. First, the built-in function ZAP of the GRAMS software package, is used to make a preliminary data selection. That portion of the spectrum is copied as x- y-data and transferred into an EXCEL worksheet and subsequently narrowed to the working range. It is desirable but not necessary to compare spectral regions having the same number of data points.

The intensity data are baseline corrected by subtracting the minimum value of the intensity:

$$Y_{i-BASE} = Y_i - \min Y_i. \quad (1)$$

These baseline corrected data are then normalized to relative intensities, NRI, using the maximum value of the baseline corrected data:

$$NRI = Y_{i-BASE}/\max(Y_{i-BASE}). \quad (2)$$

The spectra are centered on the maximum value of NRI for each compound. This x-axis transposition is implemented by subtracting the *m/z* value associated with the maximum of the NRI, $X_{@max}$:

$$X_{Rel} = X_i - X_{@max}. \quad (3)$$

The cumulative distribution funtion, CDF, is calculated from the NRI data by progressive summing of the intensities:

$$\begin{gathered} \ldots \\ CDF_3 = NRI_1 + NRI_2 + NRI_3 \\ CDF_4 = NRI_1 + NRI_2 + NRI_3 + NRI_4 \\ \ldots \\ CDF_n = NRI_1 + NRI_2 + NRI_3 + \ldots NRI_{n-1} + NRI_n. \end{gathered} \quad (4)$$

The normalized cumulative distribution function, NCDF, is formed by dividing each term of the CDF by the maximum value of the CDF, which is always the final term of the series, CDF_n:

$$NCDF_i = CDF_i/CDF_n. \quad (5)$$

This procedure leads to a function varying between zero and one. The difference between corresponding values of two NCDF functions is calculated:

$$\Delta NCDF_i = NCDF_i(A) - NCDF_i(B) \quad (6)$$

and the maximum of the absolute value, Δ_{max}, of this difference is determined.

The significance of Δ_{max} is assessed by calculating critical values of the Kolmogorov-Smirnov statistic for large n using standard methods [10]. If Δ_{max} exceeds the value of the K–S statistic at a given level of probability, then H_0 is rejected. Critical values of the K–S statistic are computed using the expression:

$$p(\alpha) = k\{(n_1+n_2)/n_1n_2\}^{0.5} \quad (7)$$

where $p(\alpha)$ is the critical value at a level of significance, α, k is a constant depending on α, and n_i is the number of data points in each of the two distributions being compared. Note that although an equal number of data points is convenient, it is not required. Values of k for differing values of α are found in standard tables [10].

Under circumstances where the values of Δ_{max} are much larger than the critical value, $p(0.001)$, we have sometimes observed that the MH^+ peaks appear to be asymmetrical. In order to evaluate the asymmetry, after determining that the shapes differ by the standard of the K–S statistic, an analysis of skewness was developed. Our analysis of skewness (Sk) is based on the difference in mass between the point at which 50% of the maximum intensity of a peak is reached relative to its centroid, m_{50}:

$$Sk = \Delta(HWHM - m_{50}) \quad (8)$$

where HWHM is half of the full-width at half maximum. If a peak is symmetrical about its centroid, then Sk should equal zero.

This approach has been implemented using macros in Excel. Copies of these macros in an Excel workbook will be supplied by the corresponding author upon request.

Results and Discussion

The validity of using smoothed versus unsmoothed data as well as establishing a basis for selecting the width of a mass window to be used in shape analysis have been evaluated. A spectrum of bovine serum albumin, BSA, a 66 kDa protein commonly used as a mass standard was chosen for the comparison. To evaluate the validity of using smoothed versus unsmoothed data, we selected a 4 kDa mass region, 64–68 kDa, centered on the MH^+ region of the spectrum. The K–S test was used to compare smoothed and unsmoothed versions of the same spectrum with NCDF functions formed from each as described above. Each version of the spectrum consisted of 707 points. The calculated maximum difference in the NCDFs, Δ_{max}, equaled 0.0412, a value less than the critical value of 0.0723 ($n = 707$) for $p = 0.05$. Thus the null hypothesis was accepted; the smoothed and unsmoothed spectra are indistinguishable, $p = 0.05$. We carried out similar comparisons of smoothed and unsmoothed data for both the tubulin and glycoconjugate materials presented below. Smoothed and unsmoothed spetra were indistinguishable in all cases. In addition, the comparison of recombinant α-tubulin with a 1:1 mixture of recombinant α-tubulin and β-tubulin, described in detail below, was made using unsmoothed data. In this case the maximum difference between the two normalized cumulative distribution functions exceeded the value obtained using the smoothed data, i.e., 0.395 versus 0.183. Since the latter value is shown below to be significant, $p < 0.001$, we conclude that, while there is some loss of the ability to distinguish between two closely related distributions after using the Savitzy-Golay smoothing algorithm, this loss does not limit our approach in any meaningful fashion. We have therefore chosen to use smoothed data for all analyses in order to avoid possible accentuation of differences arising from assigning intensity values to peak intensities that were high due to artifacts, i.e., "spikes."

In order to determine the effect of using a narrow

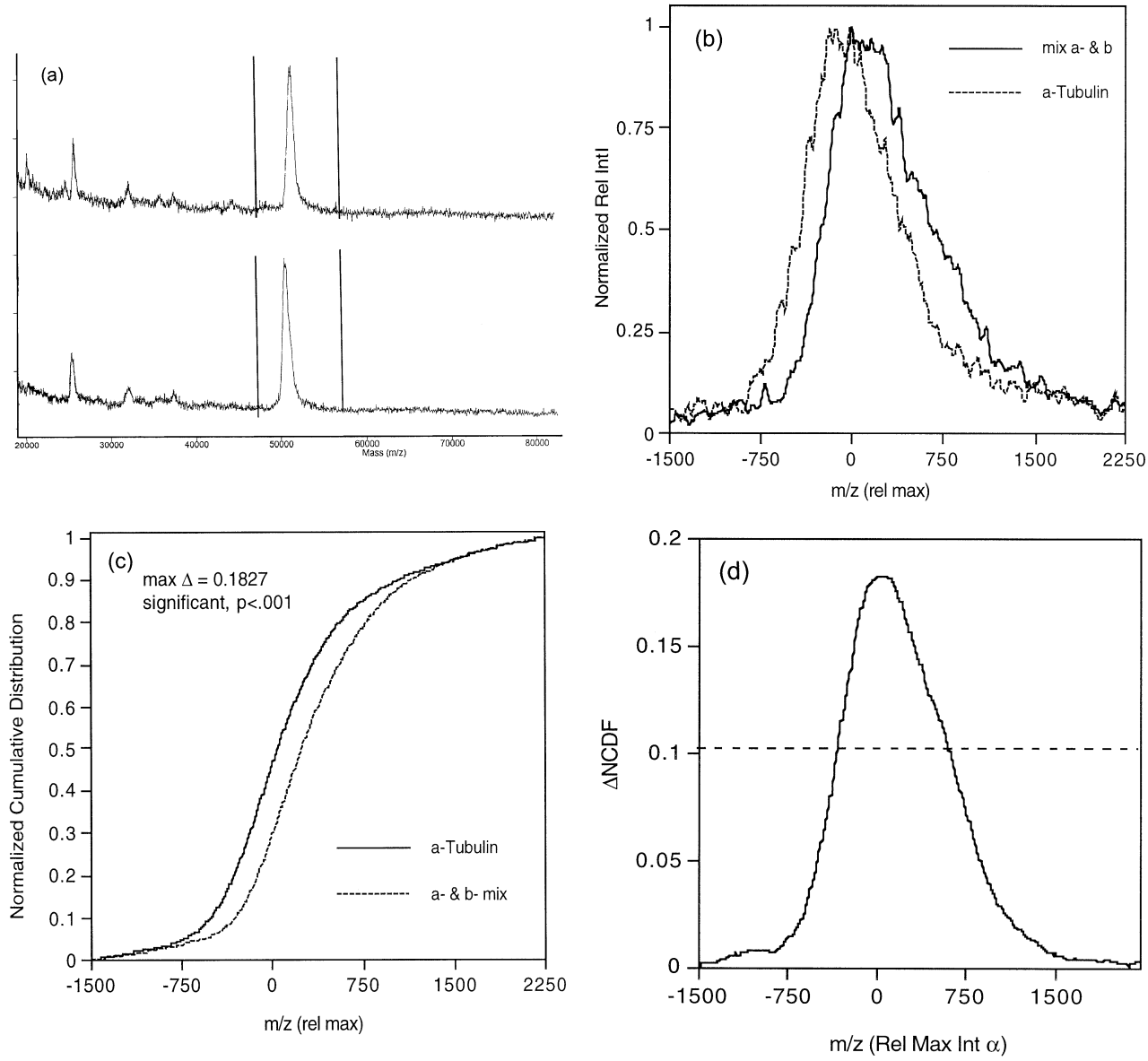

Figure 1. Progression of data manipulation steps illustrated with recombinant α- and β-tubulins, r-αTu and r-βTu. (**a**) Starting spectra of a 1:1 mixture of r-αTu and r-βTu (upper) and r-αTu (lower). Regions selected for subsequent analysis are bracketed. (*b*) Superposition of the spectra in Panel **a** following normalization to maximum intensity, NRI, and centering on the mass number of the intensity maximum, X_{Rel}. (*c*) Normalized cumulative distribution functions, NCDF, formed from each of the spectra in Panel *b*. (*d*) Plot of the difference in the NCDF functions of Panel *c*. Horizontal line shows the critical value of the K–S statistic for $p = 0.001$.

mass window for shape comparisons, a 25 kDa region of the same smoothed and unsmoothed spectra of BSA were examined. As in the previous BSA comparison, Δ_{max} was 0.0412, but the critical value of the K–S statistic differed as a consequence of the larger number of data points. For $n = 4323$, $p(.05) = 0.0351$, therefore, the null hypothesis was rejected. However, to conclude that these two spectra differed would be absurd. In this case the large value of the K–S statistic is a consequence of the substantial differences, mostly baseline, in the smoothed versus unsmoothed baselines. This example illustrates the need for keeping mass windows used for shape analysis small, but not so small as to eliminate obvious portions of the spectra being compared.

The goal of one of our collaborations is the characterization of tubulins isolated by gel electrophoresis. As part of this work it is important to be able to assess the presence of mixtures in various samples. The MALDI mass spectra of recombinant α-tubulin, r-αTu, and recombinant β-tubulin, r-βTu, were distinguishable from one another on the basis of mass assignments of their centroids. However, the MALDI spectra of a 1:1 mixture of r-αTu and r-βTu was not clearly distinguishable from that of the β-tubulin alone. Initially we

Table 1. K/S statistical analysis of BSA and tubulins

	n	Δ_{max}	$p(\alpha)$	
BSA (64–68 kDa)	707	.0412	.0732 (.05)	NS
BSA (53–78 kDa)	4323	.0412	.0351 (.05)	$p < .05$
r-αTu vs. r-βTu	741/746	.0680	.0705 (.050)	NS
r-βTu vs. 1:1 rα & βTu	746	.1797	.1010 (.001)	$p < .001$
r-αTu vs. 1:1 rα & βTu	741/746	.1827	.1011 (.001)	$p < .001$

attempted to distinguish between these two spectra using ratios of successive moments analysis. Li and co-workers [11] have successfully used this approach in the characterization of MALDI spectra of a series of synthetic polymers. Successive moments analysis of these spectra was totally ineffective in disinguishing between the MALDI spectra of tubulins, that is, the ratios of successive moments, i.e., first to second and second to third, never differed from unity. We did not attempt to use successive moments analysis in subsequent studies, but instead applied the Kolmogorov-Smirnov non-parametric statistical approach to distinguish between spectra.

The progression of data manipulation steps for the analysis of r-βTu, and the mixture of r-αTu and r-βTu is illustrated in Figure 1. Figure 1a shows the starting spectra and the regions selected by the ZAP function. Figure 1b shows the normalized, centered spectra, NRI versus X_{Rel} and Figure 1c shows the plots of NCDF versus X_{Rel} for these two spectra. Calculation of ΔNCDF for the two tublin spectra yields a $\Delta_{max} = 0.1797$. For the comparison of r-βTu and the 1:1 mixture of r-αTu and r-βTu, $p(0.001, n_1, n_2 = 746) = 0.1010$ and therefore, H_0 is rejected, $p < 0.001$. Figure 1d shows a plot of ΔNCDF versus X_{Rel} with a horizontal line denoting the critical value of the K–S statistic, $p = 0.001$. A similar comparison of r-αTu with the 1:1 mixture of r-αTu and r-βTu led to the expected ability to distinguish the two spectra, $\Delta_{max} = 0.1827$ with $p(0.001, n_1 = 741, n_2 = 746) = 0.0111$, while a comparsion of r-αTu and r-βTu as separate compounds found no differences between the two spectra, $\Delta_{max} = 0.0680$ with $p(0.05, n_1 = 741, n_2 = 746) = 0.0705$. These results are summarized in Table 1.

Another collaborative project provided the opportunity to explore the extent to which mixtures of compounds that were otherwise unresolvable might be distinguished. As part of a continuing effort to recover analytical quantities of intact proteins from electrophoretic gels [12, 13], fluorescently tagged proteins were prepared and mixed with their unlabeled form in various ratios. In addition to comparing native and labeled forms to each other, each form was compared to mixtures in the ratios 4:1, 1:1, and 1:4, native:labeled. The spectra of native ovalbumin and a 1:1 mixture of native and labeled ovalbumins are shown in Figure 2. While the masses assigned to these spectra show a small

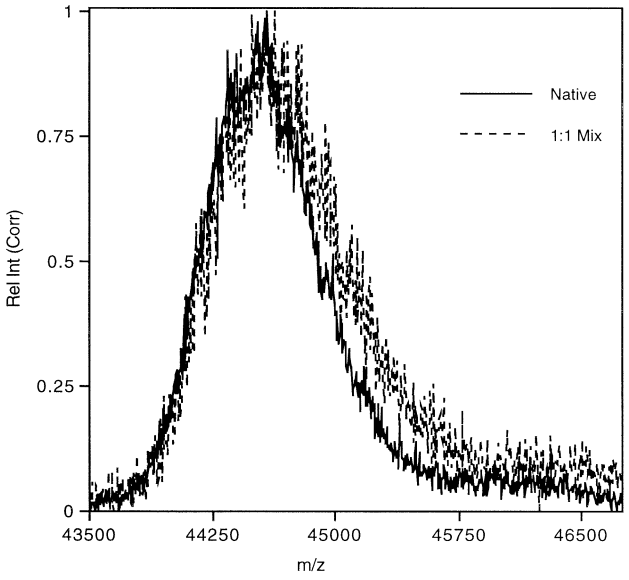

Figure 2. Superposition of spectra of native ovalbumin and a 1:1 mixture of ovalbumin and labeled ovalbumin after normalization, NRI, and centering of mass axes, X_{Rel}.

difference, no appreciable overall differences are apparent. The results of the comparisons are summarized in Table 2.

Table 2 shows that the shape of MH^+ ions of the native and labeled proteins differ significantly from one another, $p < 0.001$. Interestingly though, a mixture of these two materials at the level of 4:1 native:labeled cannot be distinguished from the native molecule. When the other two mixtures were compared with the native molecule, both were distinguishable from the native material, $p < 0.001$. On the other hand, when comparisons were made between the labeled protein and the three different mixture ratios, only the 4:1 native:labeled could be distinguished from the labeled molecule. The shapes of the MH^+ region of the spectra of the other two mixtures do not differ from the labeled molecule at at level of $p = 0.05$, i.e., the differences are not significant. The final comparison shown in Table 2 is that of the three mixtures with one another. The 4:1 mixture is clearly distinguishable form the other two mixtures, $p < 0.001$, but the 1:1 mixture cannot be distinguished from the 1:4 mixture at a level of $p = 0.05$.

Table 2. K/S statistical analysis of ovalbumin

	n	Δ_{max}	$p(\alpha)$	
Native vs. labelled	685/685	.1885	.1054 (.001)	$p < .001$
Native vs. 4:1	685/685	.0219	.0735 (.05)	NS
Native vs. 1:1	685/685	.1218	.1054 (.001)	$p < .001$
Native vs. 1:4	685/685	.1816	.1054 (.001)	$p < .001$
Labelled vs. 4:1	685/685	.1774	.1054 (.001)	$p < .001$
Labelled vs. 1:1	685/685	.0689	.0735 (.05)	NS
Labelled vs. 1:4	685/685	.0171	.0735 (.05)	NS
4:1 vs. 1:1	685/685	.1134	.1054 (.001)	$p < .001$
4:1 vs. 1:4	685/685	.1735	.1054 (.001)	$p < .001$
1:1 vs. 1:4	685/685	.0612	.0735 (.05)	NS

Table 3. Δ_{max} Calculations for four CSA glycoconjugates

	ms (kDa)	A	B	C	D
A	60	–	0.319*	0.425	0.564
B	61	–	–	0.115	0.308
C	61	–	–	–	0.225
D	62	–	–	–	–

*Critical value of K/S statistic for rejecting H_0: $p(0.001, n_1 = n_2 = 2159) = 0.0594$.

We interpret these statistical conclusions in the light of the probable heterogeneity introduced into the ovalbumin molecule by the derivatization process. At the molar ratios used for derivatization, the coupling reaction between the flourescent moiety and free amines in the native molecule is not complete. Under these circumstances the labeled molecule must exist as a mixture yielding an apprecible spreading in the isotope cluster envelope of the MH^+ ion. Inspection of the Δ_{max} values of Table 2 suggests that the labeled protein is in fact a mixture closely resembling the artificial mixture of 1:4, that is, the smallest value of Δ_{max} in Table 2 is associated with the comparison of the labeled and the 1:4 mixture.

The significance of these observations is that despite the apparent absence of differences in molecular weight or shape of the MH^+ region of the spectra, the K–S statistic allows one to distinguish mixtures from pure proteins in some circumstances. The ability to make such distinctions will certainly vary with the molecular weight of the modified protein, the resolving power of the mass spectrometer used, the molecular weight of the modifing agent, and the degree of modification. Nevertheless, the data in Table 2 demonstrate this ability for MALDI spectra for a 45 kDa protein obtained with a resolution of about 800 and modified to about 25% by an agent of molecular weight of approximately 200, a mass corresponding to about 0.4% of the weight of the native protein.

In a theoretical demonstration of this ability, we calculated the isotope cluster of myoglobin, 16,950 Da, at a resolving power of 1000, the level at which our MALDI instrument performs in the 16 kDa region, using the IsoPro [7] isotope cluster calculator. A modified form of myoglobin was created by adding 18 Da to each mass in the calculated isotope cluster. A series of mixtures of these two molecules was simulated, and 10% random noise added to the summed isotope clusters to represent typical performance in real spectra. Under these circumstances, we were able to differentiate between the native molecule and a mixture containing only 10% of the molecule that was 18 Da higher, i.e., the comaprison of native and a 9:1 mixture differed significantly, $p < 0.001$.

As a final demonstration of the utility of the K–S statistical approach, we applied the method to a series of synthetic glycoconjugates. This collaboration is part of an ongoing effort to generate synthetic vaccines [14].

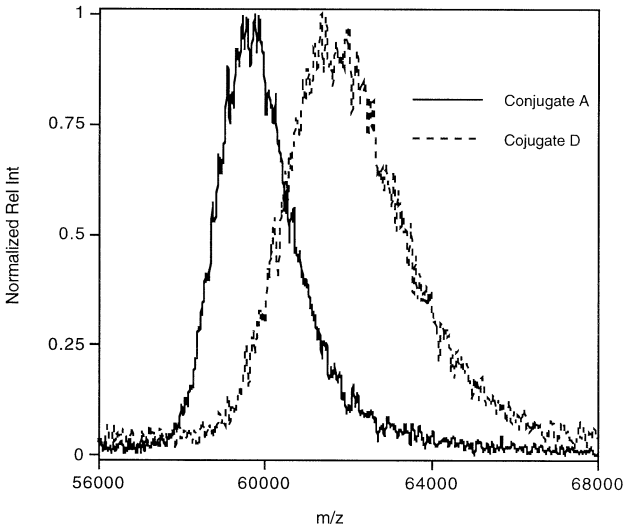

Figure 3. Superposition of spectra of synthetic Glycoconjugates A and D after normalization, NRI, and centering of mass axes, X_{Rel}.

In general, this work requires simply monitoring the extent of synthesis. However, we hypothesized that additional information could be extracted from the MALDI spectra of these materials. The data shown in Table 3 are comparisons of a series of synthetic glycoconjugates of chicken serum albumin (CSA) where the conjugating glyco residue has a molecular weight of 440. The four samples, A–D, are associated with reactions carried out over progressively longer times resulting in increasing molecular weights, as shown in the first column. Smoothed spectra of Samples A and D are shown in Figure 3. The balance of Table 3 shows the intercomparisons of each of the four samples with one another. First, it is clear that all of the entries can be differentiated from one another, $p \ll 0.001$. Second, it is clear that as the reaction progresses, the spectra relative to Sample A, for example, become more easily differentiated. This is shown graphically in the plot of ΔNCDF (A versus B, C, D) versus X_{Rel} in Figure 4 where the horizontal line denotes the critical value of the K–S statistic, $p = 0.001$.

Having established that the four CSA glycoconjugate spectra differ significantly from one another, we investigated the symmetry of the peaks. Table 4 summarizes these calculations. The basis of this calculation is that a symmetrical peak will have a skewness of zero, i.e., Sk = 0. This Table shows that for the MH^+ ion of bovine serum albumin, BSA, Sk = 3 , i.e., it is symmetrical. From the work to date it is not clear whether this value is really equal to zero within experimental error or arises from the asymmetry of the natural isotope cluster. Table 4 also shows that the four conjugates are appreciably less symmetrical than a natural protein of similar molecular weight and type, i.e., BSA. In addition, the conjugate peaks become more symmetrical as a function of reaction duration. The symmetry analysis indicated that at the shortest reaction times the peaks

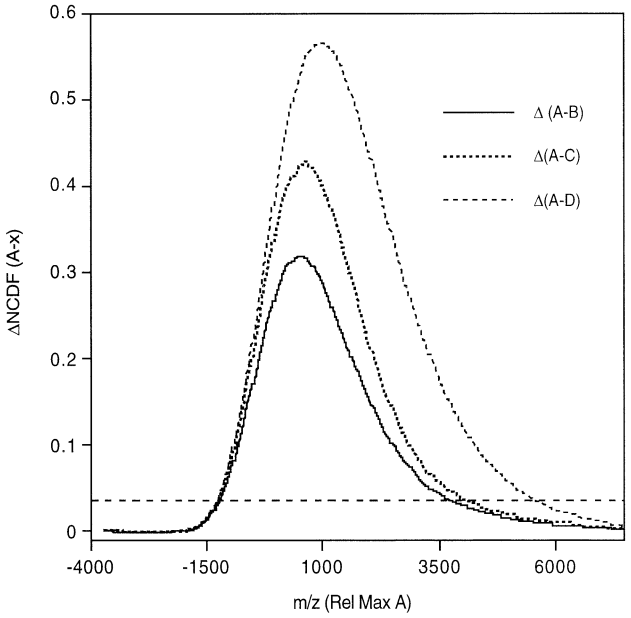

Figure 4. Plot of the difference in the NCDF functions of each of the glycoconjugates relative to Conjugate A. Horizontal line shows the critical value of the K–S statistic for $p = 0.001$.

are asymmetrical to the extent of the mass of a single glycoconjugate unit. Since the mass of the Conjugate A MH^+ ion corresponds to the addition of about 5 glycoconjugate units, the asymmetry determined corresponds to a major increased contribution to the peak shape from the presence of four units and a reduction in the contribution from six units. As the reaction proceeds, this effect is diminished so that by the time required to form Conjugate D, the MH^+ ion corresponds to the addition of an average of about 10 glycoconjugate units, but the displacement from the maximum intensity is only about 0.5 glycoconjugate units.

Summary

We have shown that it is possible to distinguish mixtures of closely related proteins from pure specimens of the same materials. We have demonstrated that in the 45 kDa mass range, mixtures of pure and modified proteins can be distinguished from one another at levels of about 25% contamination. Finally, we have shown that a combination K–S statistical and skewness analyses can be used to characterize the extent of some protein modification reactions. We argue that the K–S statistical analysis shows great promise for assessing protein heterogeneity and for detecting otherwise indistinguishable differences in distributions of molecular ions.

Table 4. Skewness of CSA conjugates

	HWHM (Da)	m_{50}	Sk
A	1005	1411	406
B	1232	1577	345
C	1165	1452	287
D	1578	1807	229
BSA	442	445	3

References

1. Creighton, T. *Structures and Molecular Properties.* W. H. Freeman and Co.: New York, 1984, pp 31–34.
2. Clauser, K., et al. Rapid Mass Spectrometric Peptide Sequencing and Mass Matching for Characterization of Human Melanoma Proteins Isolated by Two-Dimensional PAGE. *Proc. Natl. Acad. Sci. U.S.A.* **1995,** *92,* 5072–5076.
3. Yates, J. R., III, et al. Search of Sequence Databases with Uninterpreted High-Energy Collision-Induced Dossociation Spectra of Peptides. *J. Am. Soc. Mass Spectrom.* **1996,** *7(11),* 1089–1098.
4. Eng, J. K.; McCormack, A. L.; Yates, J. R., III. An Approach to Correlate Tandem Mass Spectral Data of Peptides with Amino Acid Sequences in a Protein Database. *J. Am. Soc. Mass Spectrom.* **1994,** *5(11),* 976–989.
5. Patterson, S. D.; Aebersold, R. Mass Spectrometric Approaches for the Identification of Gel-Separated Proteins. *Electrophoresis* **1995,** *16(10),* 1791–1814.
6. Figeys, D.; Aebersold, R. High Sensitivity Identification of Proteins by Electrospray Ionization Tandem Mass Spectrometry: Initial Comparison Between an Ion Trap Mass Spectrometer and a Triple Quadrupole Mass Spectrometer. *Electrophoresis* **1997,** *18(3–4),* 360–368.
7. Yergey, J. General Approach to Calculating Isotopic Distributions for Mass Spectrometry. *Int. J. Mass Spectrom. Ion Processes* **1983,** *52,* 337–349.
8. Feller, W. On the Kolmogorov-Smirnov Limit Theorems for Empirical Distributions. *Annals Math. Stat.* **1948,** *19,* 177–189.
9. Hollander, M.; Wolfe, D. *Nonparametric Statistical Methods.* John Wiley & Sons: New York, 1999, pp 178–186.
10. Beyer, W. *CRC Standard Probability and Statistics, Tables, and Formulae.* CRC Press: Boca Raton, 1991, p 337.
11. Zhu, H.; Yalcin, T.; Li, L. Analysis of the Accuracy of Determining Average Molecular Weights of Narrow Polydispersity Polymers by Matrix-Assosted Laser Desorption Ionization Time-of-Flight Mass Spectrometry. *J. Am. Soc. Mass Spectrom.* **1998,** *9,* 275–281.
12. Yefimov, S., et al. Recovery of Sodium Dodecyl Sulfate-Proteins from Gel Electrophoretic Bands in a Single Electroelution Step for Mass Spectrometric Analysis. *Analytical Chemistry* **2000,** *284,* 288–295.
13. Yefimov, S., et al. Stacking of Unlabeled Sodium Dodecyl Sulfate-Proteins with Fluorimetrically Detected Moving Boundary, Electroelution and Mass Spectrometric Identification. *Electrophoresis* **2001,** *22,* 999–1003.
14. Zhang, J., et al. Studies Toward Neoglycoconjugates from the Monosaccharide Determinant of Vibrio cholerae O:1, Serotype Ogawa Using the Diethyl Squarate Reagent. *Carbohyd. Res.* **1998,** *313,* 15–20.