Comparing Similar Spectra: From Similarity Index to Spectral Contrast Angle

Katty X. Wan, Ilan Vidavsky, and Michael L. Gross
Department of Chemistry, Washington University, St. Louis, Missouri, USA

We investigated a spectral-contrast-angle (θ) method to determine whether mass spectra of structural isomers are the same or significantly different. This method represents collisionally activated dissociation (CAD) spectra as vectors in space. Mass spectra of different isomers are represented as different vectors, having characteristic lengths and direction. The derived spectral contrast angle, which is a measure of the angle between two vectors corresponding to two closely related spectra, is a measure of whether the mass spectra are the same or significantly different. We compare this method with the similarity index (SI) method and show that the spectral contrast angle method is superior and can differentiate between very similar spectra in cases where the SI cannot. Both methods can be implemented simply in situations where the analyst is called on to decide, on the basis of mass or product-ion spectra, whether reference and unknown compounds are the same or to evaluate the reproducibility of spectra comprised of many peaks. (J Am Soc Mass Spectrom 2002, 13, 85–88) © 2002 American Society for Mass Spectrometry

Comparing and contrasting spectra are constant demands for analytical chemists, who must identify unknowns, test the reproducibility of instruments, and check the reliability of methods that are under development. Tandem mass spectrometric methods often distinguish structural isomers because isomers give characteristic product-ion spectra (i.e., fingerprints). Even for subtly different isomers that give nearly identical spectra, distinctions are difficult but still possible. The need to make judgments about similarity of spectra is growing rapidly owing to the large number of samples that can be generated from combinatorial libraries or from metabolic profiling studies. Any method that satisfies the need should not require human intervention and should be amenable to automated data processing.

To deal quantitatively with situations in which product-ion mass spectra are nearly identical, we previously developed a similarity index (SI) for comparing spectra [1]. To test for similarity of CAD or other mass spectra, the differences in signal intensities, $(i - i_0)$, at a given mass for two compounds are divided by the smaller intensity (i_0). The quotients are treated according to eq 1, where N is the number of product-ion signals that are to be used in the comparison.

$$\mathrm{SI} = \sqrt{\frac{\sum_i \left\{\frac{i - i_0}{i_0} \times 100\right\}^2}{N}} \quad (1)$$

The reproducibility of a spectrum can also be determined by comparing two spectra of the same compound taken at different times. When comparing two compounds, the differences between the spectra must be greater than the SI that was determined in the reproducibility study. Then the spectra, and their corresponding ion products or neutral structures, may be judged to be different or exist as different mixtures. Indistinguishable spectra give SI values that are equal to zero within the precision of the method.

Selecting i and i_0 so that i_0 is always smaller than i and different from zero requires human intervention. These requirements can be conflicting. If we arbitrarily designate i as the relative intensity for a peak for one isomer, A, and i_0 for another, B, then the calculated $\mathrm{SI}_{\mathrm{AB}}$ will have a different value than that of $\mathrm{SI}_{\mathrm{BA}}$ when the designations of A and B are reversed. We, therefore, suggest in accordance with reference [12] that a revised form of SI, eq 2, be used so that the arbitrary designation of i and i_0 will not give different SI values for the same pair of compounds compared.

$$\mathrm{SI} = \sqrt{\frac{\sum \left\{\frac{i - i_0}{i + i_0} \times 100\right\}^2}{N}} \quad (2)$$

In this application note, we report a simple improve-

Published online November 27, 2001

Address reprint requests to Dr. M. L. Gross, Department of Chemistry, Washington University, One Brookings Drive Campus Box 1134, St. Louis, MO 63130, USA. E-mail: mgross@wuchem.wustl.edu

1044-0305/02/$20.00
PII S1044-0305(01)00327-0

Received July 27, 2001
Revised September 17, 2001
Accepted September 17, 2001

ment of the similarity index and contrast it with another method for spectral comparison based on vector representation, whereby a spectrum is represented as a vector in an *N*-dimensional space. Product-ion spectra, for example, can then be compared based on the derived spectral contrast angles (angles between vectors). Spectral contrast is used in many applications including voice recognition [2], astronomical satellite and aerial imaging [3, 4], peak-purity determination in HPLC-UV [5], and in identification of oligonucleotides and peptides by HPLC-UV [6, 7]. In mass spectrometry, spectral contrast is used in computer library searches of unknown EI spectra [8] and for peptide sequencing using tandem spectra and protein databases [9]. In fact, spectral angle contrast (dot-product cosine) is the most reliable comparison method in library searching for compound identification [10]. Spectral angle contrast can also enhance the S/N of ion chromatograms by correlating each scan to that of a target compound [11].

The contrast angle is sufficiently simple that it can be implemented on a spread sheet without the need for special software, spectral databases, or libraries. We used in this study artificially generated spectra and real spectra from a set of isomeric oligodeoxynucleotides.

Table 1. Low energy product-ion spectra for d(TGTTT), d(TTGTT) and d(TTTGT), $[M - H]^-$ *m/z* 1482.3, and outcome of method comparisons

m/z	d(TGTTT) A[a]	d(TTGTT) B[a]	d(TTTGT) C[a]
321	24.7	22.8	43.4
625	58.5	100	0
650	0	0	58.3
705	7.4	42.9	18.6
849	23.1	0	0
874	0	34.06	35.5
929	99.3	0	4.9
954	0	41.1	53.7
1009	0	0	100
1034	30.8	23.8	0
1178	97.5	80.3	85.3
Self Similarity Index SI[b]		5.4 ± 5.4	
Self spectral angle contrast θ[b]		2.7° ± 1.6°	
	A vs. B	A vs. C	B vs. C
SI[c]	64.7 ± 1.0	85.8 ± 1.0	69.3 ± 0.4
θ[c]	50.1° ± 7.8°	67.7° ± 4.5°	61.0° ± 5.7°
SI ratio	12.0 ± 1.0	15.9 ± 1.0	12.9 ± 1.0
θ ratio	18.5 ± 0.8	25.1 ± 0.7	22.6 ± 0.7

[a]Average of 8 measured spectra.
[b]Average of 28 calculations.
[c]Average of 56 calculations.

Experimental

Materials

All deoxyoligonucleotide samples used in this study were synthesized (on the 0.2 μmol scale) at the Nucleic Acid Chemistry Laboratory at Washington University (St. Louis, MO) and were used without further purification. Samples were dissolved in 50/50 (vol/vol) MeOH/H_2O to make the final concentration approximately 20 pmol/μL.

ESI/MS/MS Experiments

Tandem mass spectrometry experiments were conducted on electrospray-produced ions by using a Finnigan LCQ instrument (San Jose, CA). The spray voltage was kept at 4.6 kV, and the capillary temperature was 200 °C. The capillary voltage was adjusted to −13.0 V for negative-ion detection. In all experiments, helium was introduced at an estimated pressure of 0.1 Pa for improving the trapping efficiency (2.6×10^{-3} Pa, indicated). The added helium also served as collision gas during CAD events. The collision energy was approximately 40% of the maximum available tickling voltage (5 V) for singly charged precursors. Data were collected and averaged for approximately 30 scans (10 groups of 3 scans each) and tabulated by using ICIS software (Finnigan, San Jose).

Artificial Spectra

A random artificial spectrum was generated on a spreadsheet. A second one was generated by small random changes of the first one. Adding a small normally distributed scatter factor such that the intensity of each peak varied by approximately 5% peak generated eight additional spectra. Thus, two nine spectra sets were available for comparison.

Processing

All the spectra measured and generated were processed using a Lotus 123 spreadsheet Version 9, using the native functions. Each separate spectrum was compared with the others and the results were averaged and analyzed. The spreadsheet files can be obtained from the authors.

Results and Discussion

We chose ESI-produced singly charged anions of isomeric pentadeoxynucleotides as a test case and obtained their low energy CAD product-ion spectra. Each isomeric compound gives a distinctive product-ion spectrum (see Table 1). We then measured in eight spectra the relative peak intensities for the compounds, designated as A, B, and C, at *m/z* 321, 625, 650, 705, 849, 874, 929, 954, 1009, 1034, 1178 (11 most intense peaks). Because the compounds are easily distinguished, we challenged the methods by generating nine very similar but artificial spectra.

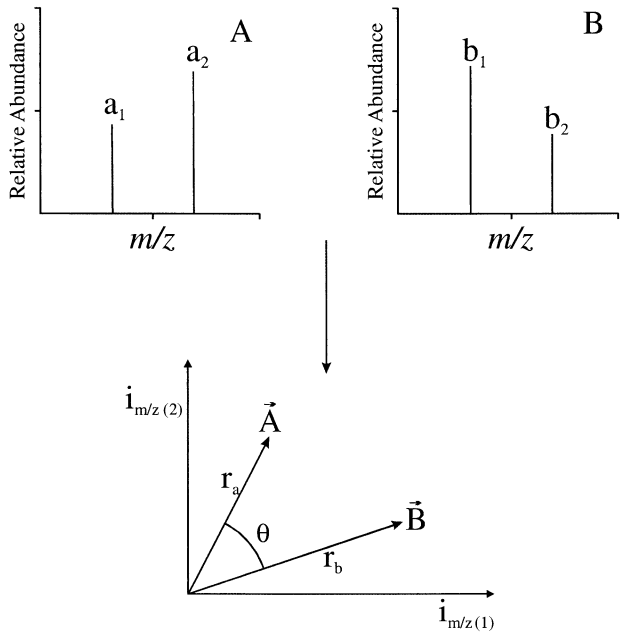

Figure 1. Schematic vector representations of the Spectra A and B.

Comparing Tandem Mass Spectra with Spectral Contrast Angle

Taking a lead from the spectral contrast angle methods used in HPLC-UV spectroscopy, we implemented vector based representations of tandem mass spectra. Figure 1 illustrates the representation for two isomeric compounds (A and B) if their product-ion spectra are presented as vectors. We show only two peaks in the scheme for simplicity, but the whole *m/z* range in the product-ion spectra can be used in the vector representation. An *N*-dimensional vector is then constructed when *N* different *m/z* values are used.

The length and direction in space of the vector is determined by the peak *m/z* and intensities. The lengths (*r*) of vectors A and B (Figure 1) are determined by eqs 3 and 4:

$$r_a = \sqrt{\sum_i a_i^2} \tag{3}$$

$$r_b = \sqrt{\sum_i b_i^2} \tag{4}$$

and are proportional to the compounds' concentrations. Mass spectra can be quantitatively compared by the derived spectral contrast angle (θ). The angle, θ, is defined as:

$$\cos\theta = \frac{\sum_i a_i b_i}{\sqrt{\sum_i a_i^2 \sum_i b_i^2}} \tag{5}$$

where a_i and b_i are the relative intensities of product-ion peaks at *m/z* value *i* for isomers A and B. An angle of zero degrees means there are no discernible spectral differences. Spectra that resemble each other have vectors that point in the same direction in the space. A 90° angle indicates a maximal spectral differentiation. Table 1 shows the calculated spectral contrast angles for comparison of the three oligodeoxynucleotide isomers.

Comparison of Spectral Contrast Angle and Similarity Index

To use the spectral contrast angle (θ) and the corrected SI (eq 2)to measure the difference in the mass spectra of isomers, we first had to determine the reproducibility of the values derived. As an example of nonsimilar spectra, we chose to measure the product-ion spectra of three oligonucleotides isomers (TGTTT, TTGTT, TTTGT denoted A, B, C respectively). Each spectrum was measured eight times on the same instrument, using the same experimental conditions. The self θ and SI were calculated for all combinations of the three product-ion spectra (Table 1). The resulting self θ was 2.7° ± 1.6° (2σ error was used), and the resulting self SI was 5.4 ± 5.4. The θ and SI were then calculated for all combinations involving the three isomers and the ratios taken for SI (second last row of Table 1) and for θ (last row of Table 1). Both the spectral contrast angle θ and the SI values are significantly higher then the self-measurements (background), and the product-ion spectra are judged to be significantly different. The calculated θ ratios are modestly but consistently higher than the SI values. The corresponding uncertainties of the derived spectral angle θ are smaller than those of the SI.

To investigate the ability to differentiate between very similar spectra, two artificial spectra were generated as described in the experimental section (see Table 2). Eight additional spectra were generated from each spectrum with small random normal distribution scatter of the intensities. The θ and SI were calculated for the two sets of nine spectra for all combinations in a similar fashion as for the oligonucleotides spectra (see Table 2). The resulting average self θ is 2.2° ± 1.0° whereas the resulting average self SI is 8.4 ± 5.7.

The ratio between the isomer θ and the self θ is, for Spectra 1 versus spectra 2, 2.2 ± 0.7. The ratio between the isomer SI and the self SI is, for Spectra 1 versus Spectra 2, 1.5 ± 1.2. The spectra were statistically very different using the spectral contrast angle θ whereas they are indistinguishable according to the SI method (ratio of 1 signifies identical spectra). This result demonstrates the advantage of using the spectral contrast angle. The spectral contrast angle is still effective when most of the spectral peaks are very similar and only small differences distinguish the compounds, whereas SI conceals the effect of small differences for a large number similar spectral peaks. This is because the SI operation involves the sum of differences between two

Table 2. Artificial similar mass spectra and outcome of comparison methods

m/z	Spec 1[a]	Spec 2[a]
100	19.3	19.6
150	59.8	62.0
200	11.0	5.6
250	55.8	50.7
300	13.6	15.7
350	61.8	60.8
400	6.7	7.2
450	48.2	46.9
500	51.8	67.8
550	100	97.8
600	36.3	34.8
650	7.6	3.2
700	50.6	53.4
750	28.7	27.0
800	64.7	64.85
850	40.6	41.9
900	42.3	40.6
950	15.1	14.1
1000	28.5	27.6
1050	51.3	52.2
1100	97.9	100
1150	7.6	8.7
1200	95.4	95.2
1250	5.7	3.2
1300	83.3	86.2
1350	77.1	71.7
1400	86.4	83.1
1450	38.1	38.8
1500	35.0	37.4
1550	14.6	11.3
Self similarity index SI[b]		8.4 ± 5.7
Self spectral angle contrast θ[b]		2.2° ± 1.0°
	Spec 1 vs. Spec 2	
SI[c]	12.3 ± 6.1	
θ[c]	4.8 ± 0.9	
SI ratio	1.5 ± 1.2	
θ ratio	2.2 ± 0.7	

[a]Average of 9 spectra.
[b]Average of 36 calculations.
[c]Average of 72 calculations.

intensities [note the term of $(i - i_0)$ in eq 1] and can be viewed as an average standard deviation. The spectral contrast operation, on the other hand, involves summing the products of two intensities [note the term of $a_i b_i$ in eq 5].

Conclusions

Although the similarity index is a useful method for spectral comparison, we found that the spectral-contrast- angle method performs better and has a lower margin of error. The latter method can also be used in any other spectroscopic application where comparison of spectra is needed. Simple software can be written and incorporated into the data processing procedure, so that automated spectra comparisons can be achieved, permitting a comparison of all the spectra of components in a mass chromatogram to a that of a target compound. Alternatively, a simpler approach using a spreadsheet can be utilized in cases where the number of comparisons to make is small. The approach should be useful for high-throughput applications such as identification of metabolites and characterization of combinatorial libraries.

Acknowledgments

This work was supported by NIH Center for Research Resources (2P41RR00954) and by the 3M Corporation.

References

1. Lay, J. O., Jr.; Gross, M. L.; Zwinselman, J. J.; Nibbering, N. M. M. A Field Ionization and Collisionally Activated Dissociation/Charge Stripping Study of Some $[C_9H_{10}]^{+\cdot}$ Ions. *Org. Mass Spectrom.* **1983,** *18,* 16–21.
2. Leek, M. R.; Dorman, M. F.; Summerfield, Q. Minimum Spectral Contrast for Vowel Identification by Normal-Hearing and Hearing-Impaired Listeners. *J. Acoust. Soc. Am.* **1987,** *81,* 148–154.
3. Lucey, P. G.; Blewett, D. T.; Taylor, G. J.; Hawke, B. R. Imaging of Lunar Surface Maturity. *J. Geophys. Res. [Planets]* **2000,** *105(E8),* 20377–20386.
4. Wald, A. E.; Kaufman, Y. J.; Tanre, D.; Gao, B.-C. Daytime and Nighttime Detection of Mineral Dust Over Desert Using Infrared Spectral Contrast. *J. Geophys. Res. [Atmos.]* **1998,** *103(D24),* 32307–32313.
5. Swartz, M. E.; Brown, P. R. Use of Mathematically Enhanced Spectral Analysis and Spectral Contrast Techniques for the Liquid Chromatographic and Capillary Electrophoretic Detection and Identification of Pharmaceutical Compounds. *Chirality* **1996,** *8(1),* 67–76.
6. Warren, W. J.; Stanick, W. A.; Gorenstein, M. V.; Young, P. M. HPLC Analysis of Synthetic Oligonucleotides Using Spectral Contrast Techniques. *BioTechniques* **1995,** *18(2),* 282–287.
7. Young, P. M.; Gorenstein, M. V. Tryptic Mapping by Reversed-Phase HPLC with Photodiode-Array Detection Incorporating the Spectral-Contrast Technique. *LC-GC* **1994,** *12(11),* 832–838.
8. Maga, J. A.; Johnson, D. L.; Morini, G. Major Volatiles in Toasted Sesame Seed Oil. *J. Food Lipids* **1995,** *4(2),* 259–268.
9. Gatlin, C. L.; Eng, J. K.; Cross, S. T.; Detter, J. C. Yates, J. R., III. Automated Identification of Amino Acid Sequence Variations in Proteins by HPLC/Microspray Tandem Mass Spectrometry. *Anal. Chem.* **2000,** *72(4),* 757–763.
10. Stein, S. E.; Scott, D. R. Optimization and Testing of Mass Spectral Library Search Algorithms for Compound Identification. *J. Am. Soc. Mass Spectrom.* **1994,** *5,* 859–866.
11. Rosenthal, D; Bursey, J. T. 20th Annual Conference on Mass Spectrometry and Allied Topics; Dallas, TX 1972; Paper T4, p 419.
12. Drahos, L.; Vekey, K. Quantification of Isomeric Differences in Mass Spectra. *Rapid Commun. Mass Spectrom.* **1996,** *10,* 1309–1315.