Simultaneous Quantitation of Amino Acid Mixtures using Clustering Agents

Ryan D. Leib, Evan R. Williams
Department of Chemistry, University of California, Berkeley, CA 94720-1460, USA

Abstract
A method that uses the abundances of large clusters formed in electrospray ionization to determine the solution-phase molar fractions of amino acids in multi-component mixtures is demonstrated. For solutions containing either four or 10 amino acids, the relative abundances of protonated molecules differed from their solution-phase molar fractions by up to 30-fold and 100-fold, respectively. For the four-component mixtures, the molar fractions determined from the abundances of larger clusters consisting of 19 or more molecules were within 25% of the solution-phase molar fractions, indicating that the abundances and compositions of these clusters reflect the relative concentrations of these amino acids in solution, and that ionization and detection biases are significantly reduced. Lower accuracy was obtained for the 10-component mixtures where values determined from the cluster abundances were typically within a factor of three of their solution molar fractions. The lower accuracy of this method with the more complex mixtures may be due to specific clustering effects owing to the heterogeneity as a result of significantly different physical properties of the components, or it may be the result of lower S/N for the more heterogeneous clusters and not including the low-abundance more highly heterogeneous clusters in this analysis. Although not as accurate as using traditional standards, this clustering method may find applications when suitable standards are not readily available.

Key words: Clusters, Quantification, Quantitative analysis, Electrospray

Introduction
Mass spectrometry is widely used to identify compounds with known structures or elucidate the structures of previously unknown compounds [1–10], even those present in complex mixtures [1, 2]. Elemental composition can be obtained from exact mass measurements [2–5] and tandem mass spectrometry can provide detailed information about structure [6–10]. However, obtaining quantitative information directly from ion abundances can be more challenging due to a number of effects, including relative ionization efficiencies, matrix effects due to the presence of other molecules, and mass-dependent ion transmission and detection efficiencies. Quantitation can be especially challenging when the efficiency of the ionization method depends significantly on both molecular structure and the matrix, as is the case for both electrospray ionization (ESI) [11–14] and matrix-assisted laser desorption ionization [15–18].

In ESI, a variety of factors can affect ionization efficiency and charging [12, 13, 19–29], such as basicity of the solvent and analyte, solvent surface tension, and analyte surface activity, which can all result in preferential ionization or suppression of individual components in mixtures. Separation methods, such as liquid chromatography, can be used to reduce matrix effects and are commonly used with complex mixtures [30, 31], but increase analysis time and do not necessarily eliminate matrix effects or effects of differential ionization efficiencies [14, 32].

Accurate quantitation can be done by using carefully chosen and/or specifically prepared standards [31–44].
Typically, standards are selected to closely mimic the physical properties of the analyte, with the most robust quantitation done using isotopically labeled forms of the analyte as internal standards [38–41]. Both internal and external standards are used in a variety of analytical applications of mass spectrometry [42–44]. However, suitable standards may not always be readily available, such as with newly discovered natural products, illicit or restricted compounds, or new products and intermediates formed by organic synthesis, which can make it difficult to rapidly obtain accurate quantitation with mass spectrometry.

We recently introduced a new approach to obtaining quantitative measurements of analyte molar fractions directly from an ESI mass spectrum without using traditional standards [45–47]. A clustering agent, such as an amino acid, is added to a solution in significant molar excess, at a concentration typically around 1 to 10 mM. Abundant homogenous clusters of the added agent are formed, as are heterogeneous clusters that contain primarily the clustering agent but also one or more analyte molecules. Smaller clusters often exhibit preferential incorporation of some components, or can preferentially ionize depending on the cluster composition. But these effects become smaller with increasing cluster size where incorporation of analyte molecules into the cluster becomes more statistical and reflects the relative ratios of components in solution. The abundances of these nonspecific clusters can be used to obtain the molar fractions of the components in solution [45]. Even when serine, which is known to form specific chirally selective structures at small cluster size [48–50], is used as a clustering agent, the composition of large clusters can be used to obtain solution molar fractions of other amino acids that are accurate to within ~20% of the solution value [46]. By adding a known amount of the clustering agent to the solution, the absolute concentration of an analyte can be determined [47]. This method has been demonstrated on solutions containing individual amino acids and peptides, and has been applied to the direct analysis of active ingredients in Tamiflu and other pharmaceutical tablets, where the ionization/detection efficiency of individual components differed by up to 100-fold, but the dosages of the active ingredients in each of the tablets were determined to typically better than 20% accuracy [47]. Although not as accurate as methods that use more traditional standards, this method has the advantages that it is fast, it can be used for mixtures containing unknown analytes, and can be used when suitable standards may not be readily available, such as schedule I or II controlled substances, or designer drugs that have not been previously characterized. Here, we investigate the viability of this cluster agent approach for obtaining simultaneous quantitative information from more complex solution mixtures containing up to 10 components including either serine or tryptophan as clustering agents.

**Experimental**

**Mass Spectrometry**

All mass spectra were obtained using a 9.4 T Fourier transform ion cyclotron resonance (FT/ICR) mass spectrometer that has been described elsewhere [45, 51]. Aqueous stock solutions containing glycine, alanine, serine, threonine, leucine, lysine, histidine, phenylalanine, arginine, and tryptophan (Sigma Aldrich, St. Louis, MO, USA) were prepared at 6 mM. Mixed analyte solutions were prepared from these stock solutions and diluted to a final concentration of 3 mM. Ions were formed by nanoelectrospray ionization using ~10 μL of aqueous solution loaded into borosilicate capillaries pulled to a tip inner diameter of ~2 μm. The borosilicate capillary is positioned ~2–3 mm away from the source inlet and electrospray is initiated by applying approximately ~1 kV to the source inlet. A platinum wire in direct contact with the analyte solution in the borosilicate capillary is grounded. Ions are accumulated in an external hexapole for 1.5 s prior to injection and trapping within the ion cell. Trapping of clusters is enhanced by pulsing N2 gas through a piezoelectric valve to increase the cell pressure transiently to ~1×10−6 Torr, after which the cell pressure returns to ~1×10−9 Torr prior to ion detection. Spectra were signal averaged to increase the signal-to-noise ratio.

**Results and Discussion**

**Solution Concentrations and Cluster Abundances**

If clusters are formed statistically from a solution containing two or more components, the cluster abundances can be used to determine the relative concentrations of the components in solution if the effects of differential ionization, detection, and ion transmission are small, which should occur when one of the components is dominant within the clusters. For a two-component mixture, molar fractions can be obtained using either a binomial expansion or, more rigorously, a weighted average [45]; in cases where the absolute concentration of one component is known, the absolute concentration of the other component can be readily obtained. In the same way, cluster ion abundances from solutions containing more than two components can also be used to determine absolute molar fractions for each component. In Figure 1a and b, theoretical ion abundances for clusters containing the same number of subunits, n, are shown for two different two-analyte mixtures that share the same majority “clustering agent”, C, but one of two different minority analytes, A and B, respectively. In both cases, homogeneous clusters composed only of the clustering agent C are formed, as well as a series of heterogeneous clusters still composed of n subunits, but containing one or more analyte molecules. The mass difference between each of these clusters is the difference in the molecular weight of a single clustering agent molecule and a single analyte.
molecule, denoted by A-C and B-C in the two spectra, respectively. Multiple incorporations of an analyte are also expected if the cluster size or the analyte solution molar fraction are sufficiently large, resulting in additional clusters separated by integer values of the mass difference between the clustering agent and the analyte. To obtain a solution percent molar fraction, \( F_m \%), from either of these theoretical cluster distributions, the weighted average in Equation (1) can be used:

\[
F_m \% = \frac{\sum h \frac{I_h}{n}}{\sum h I_h} \times 100
\]  

(1)

where \( I \) is the ion abundance of each cluster, \( n \) is the number of molecules in the cluster and \( h \) is the number of minority analyte molecules incorporated into the cluster.

When clusters are formed from a solution containing more than two components, the above weighted average can still be used iteratively for each analyte to obtain solution molar fractions from the resulting distribution of clusters as long as all cluster ions at a given \( n \) containing the specific analyte are included. As an example, Figure 1c shows a theoretical cluster distribution formed for a solution containing the clustering agent C and both analytes A and B for the same number of subunits \( n \). Note that the addition of the second analyte greatly reduces the abundance of the homogeneous clustering agent peak because the clustering agent is now largely present in the heterogeneous clusters, including a cluster corresponding to the incorporation of both minority analytes, with a mass difference of \( A + B - 2C \).

Although this analysis can be continued iteratively for additional incorporations of minority components into a cluster, (e.g., \( A + A + B - 3C \), \( A + B + B - 3C \)), these higher order analyte incorporations should be highly unlikely when the clustering agent is added in large excess, e.g., in these experiments, greater than ~17-fold. An excess of clustering agent reduces the observed spectral overlap between cluster ions containing multiple heterogeneous components by shifting most of the observed ion abundance into the homogeneous cluster and heterogeneous clusters containing only one or two analyte molecules. Only the incorporation of up to two minority analyte components is considered in these analyses.

\section*{Molar Fractions for Four-Component Mixtures}

An ESI mass spectrum of a solution containing four amino acids, serine, histidine, arginine, and leucine, prepared at an approximately 95/2/2/1 ratio is shown in Figure 2. In addition to the protonated molecules, homogeneous and heterogeneous cluster ions from dimers to clusters containing up to ~35 molecules are observed. The heterogeneous clusters consist of primarily serine and one or two other analyte molecules (see inset). For this four-component mixture, 10 clusters of a given \( n \) are used to calculate relative solution phase molar fractions: the homogeneous cluster, the three heterogeneous clusters containing only one minority component, i.e., histidine, arginine, or leucine, and the six possible heterogeneous clusters containing two of the same minority component, e.g., two histidines, or two different minority components, e.g., an arginine and a leucine. Because the serine clustering agent is present in significant excess, the majority of the heterogeneous cluster ion abundance corresponds to the incorporation of a single analyte. Greater ion abundance of clusters containing multiple different analytes would be expected at either larger cluster sizes or at higher relative molar fractions. For example, 10 clusters with \( n=29 \) are observed: a homogeneous serine cluster, as well as nine heterogeneous clusters corresponding to the incorporation of up to two analyte molecules. Using the abundances of these 10 clusters in Equation (1), solution molar fractions of 1.36%, 1.78%, and 1.05% are determined for arginine, histidine, and leucine, respectively. Excluding the six heterogeneous clusters corresponding to multiple incorporations and using only the abundances of the four clusters corresponding to the homogeneous cluster and the incorporation of a single analyte molecule in Equation (1) results in solution molar...
fractions of 0.56%, 1.18%, and 0.47% for arginine, histidine, and leucine, respectively. Thus, without including the contribution of clusters containing multiple analyte molecules, the solution molar fractions obtained by this method are artificially low because the abundances of clusters containing multiple analyte molecules are significant.

Solution molar fractions for each analyte at various cluster sizes were obtained by solving Equation (1) iteratively for each analyte in the mixture and these values are shown in Figure 3. The most intense protonated molecule in the mass spectrum is not the clustering agent serine but arginine, which comprises 61% of the molecular ion abundance even though it is present at only a 2% solution molar fraction, a 30-fold excess. Histidine also ionizes efficiently, comprising 33% relative protonated molecular ion abundance, 15-fold higher than its molar fraction in solution. Interestingly, the relative ion abundance of protonated leucine is ~1%, although the similarity between relative protonated molecule abundance and solution molar fraction for this analyte is almost certainly coincidental.

Previous results for leucine-serine mixtures showed a strong enhancement in formation of protonated leucine, 54-fold in excess of its solution molar fraction [46]. As a result of the anomalously high ion abundances of arginine and histidine,
protonated serine, the primary component of the mixture, is only ~5% relative abundance, 19-fold less than its 95% solution molar fraction. The anomalously high abundances of the protonated arginine and histidine relative to the clustering agent are likely due to their high basicity, although differences in surface activity and instrumental parameters can also affect relative ion abundances.

In contrast to the protonated molecules, the cluster compositions rapidly reflect the relative solution molar fractions with increasing cluster size. Even though serine is only ~5% of the molecular ion signal, it represents 88% and 96% of the composition of all dimers and trimers, respectively. This indicates that incorporation of molecules present in the solution into the cluster ions is more statistical, although some specificity is still observed at these small sizes. For example, dimers of serine and histidine appear to be preferentially formed, comprising 11% of all dimer ions, corresponding to a ~6% molar fraction. Arginine is preferentially excluded from the trimer, comprising only 0.3% of the ion abundance, corresponding to a 0.1% molar fraction.

For the octameric clusters, histidine and arginine both incorporate at a much lower ratio than expected statistically, resulting in measured molar fractions of 0.08% and 0.01%, respectively. Leucine, however, incorporates statistically at ~1%. As has been reported previously, the serine octamer typically forms a chirally selective specific structure that has been demonstrated to exclude a number of other amino acids that disrupt the octamer structure [48–50].

For cluster ions with $n$ between 19 and 33 (Figure 3, inset), average molar fractions of 1.97%, 1.55%, and 1.02% are obtained for arginine, histidine, and leucine, respectively. At these larger cluster sizes, the compositions are largely independent of cluster size, and correlate well with the solution values of 1.93%, 1.95%, and 0.90% for these respective analytes. Thus, the composition of large, gas-phase clusters reflects the solution composition to within 25% accuracy.

To determine if the cluster compositions are sensitive to small changes in the solution composition, an ESI mass spectrum of a mixture of serine, histidine, arginine, and leucine in an approximately 95/2/1/2 ratio, respectively, was obtained, and the percent molar fractions calculated from this spectrum are shown in Figure 4. For the protonated molecules, histidine is 48% of the total ion abundance, with arginine and leucine comprising 33% and 5%, respectively, inconsistent with their 2/1/2 solution ratios. However, values obtained from large cluster ions with $n=19–28$ are 1.59%, 0.85%, and 1.41% for histidine, arginine, and leucine, respectively, reasonably consistent with their respective solution molar fractions of 1.93%, 0.98%, and 1.81%. Although nonstatistical cluster composition may contribute, the slightly lower values obtained in this experiment are likely an artifact of the low cluster ion S/N ratio. Noise disproportionately affects low signal-to-noise ratio heterogeneous peaks [45], such as those observed here, resulting in slightly lower values when analyzed by a weighted average. Improving the S/N through additional signal averaging or reducing chemical noise due to the other nonspecific adducts, such as salts, would likely improve the accuracy of these measurements. Even with these caveats, the abundances and composition of the cluster ions can be used to obtain a moderately accurate measure of solution molar fractions (~20% for histidine), whereas the abundances of

![Figure 4](image-url)
protonated molecules can differ dramatically from their solution molar fractions (30-fold for histidine). These results show that this cluster method can be used to measure small changes in relative solution concentration of the analytes.

Solution Molar Fractions from More Complex Mixtures

To investigate the extent to which this clustering agent method can be applied to more complex mixtures, two solutions were prepared. Each solution contained 10 components: a clustering agent (tryptophan; 87%) and different concentrations of nine other amino acids. The minority components glycine, alanine, serine, threonine, leucine, phenylalanine, and arginine are 1/1/1/1/1/1/1/1/1 percent and 1/1/1/1/1/1/1/1/1 percent, respectively, and ESI mass spectra of these two solutions are shown in Figure 5a and b, respectively. Tryptophan was selected as a clustering agent because its mass is roughly twice that of serine. The resulting increase in m/z spacing between each homogeneous cluster reduces spectral overlap of the many possible heterogeneous clusters that could be formed.

For the protonated molecules, significant differences between the relative abundances and the solution molar fractions are observed. For some analytes, ionization efficiency is significantly enhanced. Arginine is present at a 1% molar fraction in both solutions but the relative abundance of protonated arginine is 6% and 3%, respectively, in the ESI spectra from the two solutions. The solution molar fractions of lysine and histidine are each 5% in these respective solutions, yet their relative protonated molecule signals are both 13%. For many of the other analytes, ionization is suppressed. Both protonated alanine and protonated threonine have relative abundances of 0.02% and 0.01% in the respective solutions, corresponding to 50- and 100-fold suppressions in their ion abundances relative to their solution molar fractions. Thus, as was observed for the four-component mixtures, the relative abundances of the protonated molecules correlate poorly with solution molar fraction in these more complex mixtures.

In addition to the protonated molecules, cluster ions with n up to 22 are also formed from these solutions (Figure 5). Homogeneous cluster ions containing only tryptophan are observed, as are a host of heterogeneous cluster ions corresponding to the incorporation of one or two analyte molecules into a tryptophan cluster. The molar fractions determined from these cluster ions for each of the nine

![Figure 5. ESI mass spectra of solutions containing tryptophan, lysine, histidine, glycine, alanine, serine, threonine, leucine, phenylalanine, and arginine in differing ratios: (a) 87/5/1/1/1/1/1/1/1/1 and (b) 87/1/5/1/1/1/1/1/1/1, respectively. An expanded region of the spectrum shows homogeneous and heterogeneous clusters with selected clusters denoted (see text)](image)

![Figure 6. Percent molar fractions obtained from clusters formed from solutions containing tryptophan, lysine, histidine, glycine, alanine, serine, threonine, leucine, phenylalanine, and arginine in differing ratios: (a) 87/5/1/1/1/1/1/1/1/1 and (b) 87/1/5/1/1/1/1/1/1/1, respectively, as a function of cluster size)](image)
analytes are shown as a function of cluster size in Figure 6. At small $n$, formation of homogeneous tryptophan clusters is favorable, and most heterogeneous clusters are suppressed. However, the abundances and composition of heterogeneous clusters begins to more closely reflect the solution molar fractions at larger cluster sizes. For $n=17$ and larger, the average solution molar fraction obtained from clusters by this method is within 25% for glycine and within a factor of $\sim 2$ for most other analytes (Supplemental Table 1).

Although mole fractions obtained from the cluster data provide a significantly more reliable indication of solution composition compared with the individual protonated molecule abundances, these values are not as accurate as those obtained for less complex mixtures. There is evidence for either specific abundances, these values are not as accurate as those obtained from clusters by homogeneous tryptophan clusters. For the $n=20$ clusters, the heterogeneous peak containing a single lysine (a) or histidine (b) should be 1.15 times more abundant than the homogeneous tryptophan cluster if these cluster compositions are statistical. However, the respective relative abundances are 74% and 29% (Figure 5, insets). This indicates that incorporation of both lysine and histidine into this large tryptophan cluster is hindered, and that incorporation of histidine is less favorable than lysine.

Additional evidence for specific incorporation into tryptophan clusters is found at other cluster sizes. This suggests that clusters formed from these mixtures may not just occur as a sequential addition of individual amino acids. Large clusters could also be formed through the aggregation of smaller clusters, such as dimers and trimers. Small clusters more readily form specific structures and incorporation of these specific structures into larger clusters would skew the observed ion abundances of larger heterogeneous clusters to reflect the less statistical incorporation.

Although lysine and histidine show different extents of incorporation into the various heterogeneous clusters, the measured molar fractions obtained for lysine and histidine are essentially the same: 1.7% and 1.6%, respectively, for clusters with $n=17$ through 22. Even though specific cluster formation occurs, the sum total composition analysis over a wide range of cluster sizes more accurately reflects solution molar fraction than data for an individual cluster size. This suggests that cluster formation may occur stoichiometrically, if not statistically, at each cluster size, which is consistent with these clusters being predominately formed by a charged residue mechanism, as has been reported previously [46, 50]. Although this method is clearly not as accurate as techniques using traditional standards, this method offers a significantly more reliable indicator of solution composition than the abundances of individual protonated molecules and provides rapid, albeit rough, quantitative information even from relatively complex mixtures.

**Conclusions**

The compositions of mixtures containing either four or 10 amino acids were analyzed by using the abundances of both homogeneous and heterogeneous clusters formed by ESI. Although the relative abundances of some of the protonated molecules differed from their molar fractions in solution by as much as two orders of magnitude, the molar fractions determined from larger clusters were within 25% for the four-component solutions although poorer accuracy was obtained for the 10-component mixtures where the solution molar fractions could typically be determined within a factor of three. This indicates that the accuracy of this cluster quantitation method decreases with increasing mixture complexity, but this method can still provide some quantitative information directly from ion abundances in an ESI mass spectrum.

There are several challenges in extending this method to more complex mixtures. With increasing mixture complexity, the ion signal is spread into many additional clusters, reducing the overall signal-to-noise ratio of a given cluster. This also increases the resolving power required to separate all the different clusters. Preferential incorporation of some components into the clusters may occur for solutions that contain molecules that have vastly different physical properties. Multiple measurements using different clustering agents may reduce error associated with specific incorporation of some analytes.

Although not as accurate as traditional methods using either internal or external standards, this cluster quantitation method does have the advantages that the analytes do not need to be identified, and quantitative information for all analytes can be obtained simultaneously. This cluster quantitation method may be advantageous when combined with separations or when there is a limited number of unknown analytes, such as mixtures containing intermediates and side reaction products generated during the synthesis of organic or pharmaceutical molecules, or with illicit drugs of unknown structure.

**Acknowledgments**

The authors are grateful for financial support from the National Institutes of Health (grant R01 GM064712-08).

**Open Access**

This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

**References**

1. Domon, B., Aebersold, R.: Mass spectrometry and protein analysis. *Science* **312**, 212–217 (2006)
47. Flick, T.G., Leib, R.D., Williams, E.R.: Direct standard-free quantitation of tamiflu and other pharmaceutical tablets using clustering agents with electrospray ionization mass spectrometry. *Anal. Chem.* **82**, 1179–1182 (2010)

48. Nanita, S.C., Cooks, R.G.: Serine octamers: Cluster formation, reactions, and implications for biomolecule homochirality. *Angew. Chem. Int. Ed.* **45**, 554–569 (2006)

49. Hodyss, R., Julian, R.R., Beauchamp, J.L.: Spontaneous chiral separation in noncovalent molecular clusters. *Chirality* **13**, 703–706 (2001)

50. Spencer, E.A.C., Ly, T., Julian, R.R.: Formation of the serine octamer: Ion evaporation or charge residue? *Int. J. Mass Spectrom.* **270**, 166–172 (2008)

51. Robinson, E.W., Williams, E.R.: Multidimensional separations of ubiquitin conformers in the gas phase: relating ion cross sections to H/D exchange measurements. *J. Am. Soc. Mass Spectrom.* **16**, 1427–1437 (2005)