J|A|C|S
ARTICLES
Published on Web 08/16/2006

Mass Measurements of Increased Accuracy Resolve Heterogeneous Populations of Intact Ribosomes

Adam R. McKay, Brandon T. Ruotolo, Leopold L. Ilag,† and Carol V. Robinson*

Contribution from the Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge, CB2 1EW, U.K.

Received March 10, 2006; E-mail: cvr24@cam.ac.uk

Abstract: It is established that noncovalent complexes can be maintained both during and after electrospray and that assemblies of increasing size and complexity often lead to broadened peaks in mass spectra. This broadening arises from the tendency of large protein assemblies to form adducts with salts and is compounded when complexes are isolated directly from cells, without the full protein complement. To investigate the origins of this broadening in mass spectral peaks and to develop the optimal method for analyzing mass spectra of large protein complexes, we have carried out a systematic investigation of a series of noncovalent complexes representing a range of different sizes and architectures. We establish a positive correlation between peak width and the increased mass observed and show that this correlation is independent of the instrumental parameters employed. Using this relationship we show that we can determine masses of both 30S subunits and intact 2.3 MDa 70S ribosomes from *Thermus thermophilus*. The masses of both particles are consistent with multiple populations of ribosomes. To identify these various populations we combine simulated mass spectra of ribosomes, with and without the full protein complement, and estimate the extent of adducts from our study of known complexes. The results allow us to determine the contribution of the different subpopulations to the overall mass spectrum. We confirm the existence of these subpopulations using tandem mass spectrometry of intact 30S subunits. Overall, the results show that, rather than uniform particles, gas-phase ribosomes consist of a number of discrete populations. More generally, the results establish a rigorous procedure for accurate mass measurement and spectral analysis of heterogeneous macromolecular assemblies.

Introduction

The ability to determine the elemental composition of an organic substance from an accurate mass measurement was first demonstrated in 1954 and had a major impact on chemistry.[1] Similarly in the postgenomic era, the ability to determine accurately the mass of peptide fragments using mass spectrometry (MS) has revolutionized the field of proteomics.[2] In parallel with the development of MS for proteomics, fuelled largely by the introduction of "soft" ionization techniques such as electrospray, specialized instrumentation that allows for the analysis of labile noncovalent complexes has emerged. Specifically nanoelectrospray,[3,4] a technology originally focused on reduced sample consumption, is ideally suited to the study of large noncovalently associated protein assemblies analyzed from aqueous buffered solutions.[5] This together with instrumental advances that have improved transmission in quadrupole-time-of-flight analyzers[6] (QToF) have allowed the observation and investigation of species as large as viruses[7,8] and as complex as ribosomes.[9,10]

For noncovalent complexes, ionization using nano-electrospray produces highly charged ions from small droplets through a mechanism known as the charge residue model.[11] Successive off-spring droplets are formed after the parent droplet undergoes solvent evaporation and Coulombic fission. Continued evaporation leads to gas-phase ions retaining the charge of the last droplet. If the droplet contains soluble protein, adduct formation will produce ion−adduct complexes. We and others have shown previously that the measured mass of noncovalent assemblies is often higher than expected from the sum of their components.[12−15] We attribute this mass increase to insufficient removal of either solvent and/or buffer ions during the elec-

† Present address: Department of Analytical Chemistry, Stockholm University, Stockholm, S 106 91, Sweden.

(1) Beynon, J. H. *Nature* **1954**, *174*, 735.
(2) Henzel, W. J.; Billeci, T. M.; Stults, J. T.; Wong, S. C.; Grimley, C.; Watanabe, C. *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 5011−5015.
(3) Wilm, M.; Mann, M. *Int. J. Mass Spectrom. Ion Processes* **1994**, *136*, 167−180.
(4) Wilm, M.; Mann, M. *Anal. Chem.* **1996**, *68*, 1−8.
(5) Chung, E. W.; Henriques, D. A.; Renzoni, D.; Morton, C. J.; Mulhern, T. D.; Pitkeathly, M. C.; Ladbury, J. E.; Robinson, C. V. *Protein Sci.* **1999**, *8*, 1962−1970.
(6) Sobott, F.; Hernandez, H.; McCammon, M. G.; Tito, M. A.; Robinson, C. V. *Anal. Chem.* **2002**, *74*, 1402−1407.
(7) Tito, M. A.; Tars, K.; Valegard, K.; Hajdu, J.; Robinson, C. V. *J. Am. Chem. Soc.* **2000**, *122*, 3550−3551.
(8) Fuerstenau, S. D.; Benner, W. H.; Thomas, J. J.; Brugidou, C.; Bothner, B.; Siuzdak, G. *Angew. Chem., Int. Ed. Engl.* **2001**, *40*, 982.
(9) Benjamin, D. R.; Robinson, C. V.; Hendrick, J. P.; Hartl, F. U.; Dobson, C. M. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 7391−7395.
(10) Rostom, A. A.; Fucini, P.; Benjamin, D. R.; Juenemann, R.; Nierhaus, K. H.; Hartl, F. U.; Dobson, C. M.; Robinson, C. V. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 5185−5190.
(11) Dole, M.; Mack, L. L.; Hines, R. L.; Mobley, R. C.; Ferguson, L. D.; Alice, M. B. *J. Chem. Phys.* **1968**, *49*, 2240−2249.
(12) Nettleton, E. J.; Sunde, M.; Lai, Z.; Kelly, J. W.; Dobson, C. M.; Robinson, C. V. *J. Mol. Biol.* **1998**, *281*, 553−564.

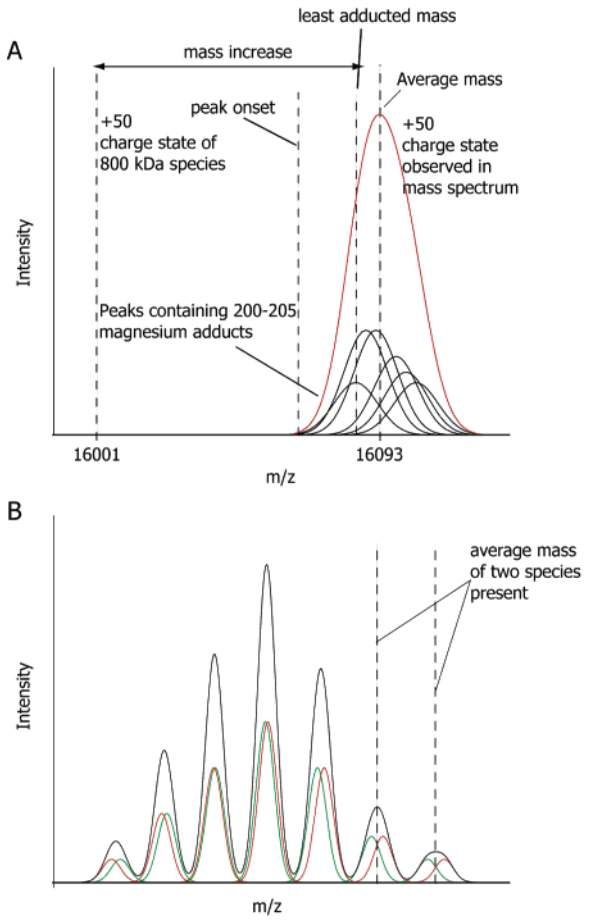

Figure 1. (A) Illustration showing the effect of sodium or magnesium adduct formation on the 50+ charge state of an 800 kDa species in an electrospray mass spectrum. The observed peak, shown in red, is comprised of a series of unresolved Na/Mg adducts. Na/Mg adducts were chosen as they are readily identified in the charge state series of monmeric proteins released from complexes in tandem MS experiments. The intensities of peaks used for this model were set to produce a Gaussian distribution of adducts found typically in electrospray spectra. It is possible to calculate an average mass of the complex by taking the *m*/*z* value at the centroid of the peak or a "least adducted mass" by estimating the *m*/*z* value near the onset of the peak. (B) Simulation of two overlapping charge state series and the resulting mass spectrum. It is clear that any attempt to measure the mass of this series using the centroid *m*/*z* will result in an average mass of the two species present.

trospray process resulting in an increase in the measured mass as well as a broadening of the peaks in the spectrum.

To explore the contribution of small molecule adducts we have simulated the mass spectrum for a hypothetical 800 kDa protein bound to a range of small cations, in this case 200−205 sodium or magnesium ions, Figure 1A. The peaks observed in the mass spectrum are a product of a series of overlapping charge states corresponding to protein complex ions that adhere to different amounts of Na^+/Mg^{2+}. These protein−adduct ions combine to produce a peak at higher molecular mass than predicted from the sequence of the hypothetical 800 kDa protein complex. Although this "noncovalent mass shift'" can be manipulated by control of the pressures within the mass spectrometer[15] and the use of a heating apparatus to aid in the desolvation process,[16] it is often impossible to yield ions that when measured have a mass comparable to that predicted by sequence. It is also important to note that there is a large potential for error in determining the *m*/*z* value of the composite charge state. For example, mass measurement at the centroid value of the peak will produce an average mass over all adducted species present in the sample (indicated as "average mass" in Figure 1A), and a mass measured at the appropriate point on the leading edge of the observed signal will result in a mass corresponding to the "least-adducted" mass of the noncovalent complex. Even though the "least-adducted" mass is often a useful measurement in cases where a small amount of adduction is observed, the correct position along the leading edge of the charge state becomes more difficult to define for assemblies that retain a larger number of buffer and solvent molecules.

An additional factor that influences the width of charge states observed in the mass spectra of large noncovalent complexes is incomplete binding of the full complement of proteins in a large assembly. A simulated charge state series for a hypothetical 800 kDa protein complex with both the full complement of proteins and one in which a 60 kDa protein is absent is shown, Figure 1B. Using a peak width at full width half-maximum (fwhm) of 250 *m*/*z*, a value typically encountered during experiments with noncovalent complexes, we simulated the two charge state series for the two protein complexes. The simulated spectrum shows that the two series are largely overlapping. Measuring the molecular mass of the product of these two charge state series may give rise to an average mass of the two complexes present. This simulation illustrates how heterogeneity due to protein composition can produce unresolved peaks, thus distorting the measured mass and rendering the composition of the noncovalent complex difficult to define.

Both the formation of adduct ions and the heterogeneity of the protein complement of the noncovalent complex can lead to peak broadening which in turn compromises our ability to determine accurately the *m*/*z* ratio of a given charge state. It is clear however that discriminating between formation of adducts and heterogeneity in the protein composition of these large noncovalent complexes will advance significantly our ability to characterize unknown protein assemblies. In this study we describe a series of experiments that allow us to estimate the extent of adduct retention after desolvation for a complex of unknown mass. This enables us to determine the protein composition of the 30S ribosomal subunit, an 800 kDa protein−RNA complex with a full complement of 20 proteins and one RNA molecule, approximately 1500 bases in length. We show that 30S subunits consist of four distinct species. Further application to intact 70S ribosomes, 2.3 MDa particles composed primarily of 3 large RNA molecules and 58 proteins, demonstrate that the heterogeneity identified in 30S subunits can also be discerned in spectra recorded for intact 70S ribosomes.

Experimental Section

Mass Spectrometry. All spectra were acquired on a Q-ToF 2 modified for high mass operation and equipped with a Z-spray nanoflow source (Waters, Manchester, UK).[6] The following experimental parameters were used: capillary voltage 1.7 kV, cone voltage 80−120 V, cone gas 100 L/h, collision cell voltage up to 200 V, ion transfer

(13) Rostom, A. A.; Robinson, C. V. *J. Am. Chem. Soc.* **1999**, *121*, 4718−4719.
(14) Green, B. N.; Gotoh, T.; Suzuki, T.; Zal, F.; Lallier, F. H.; Toulmond, A.; Vinogradov, S. N. *J. Mol. Biol.* **2001**, *309*, 553−560.
(15) Sobott, F.; Robinson, C. V. *Int. J. Mass Spectrom.* **2004**, *236*, 25−32.
(16) Benesch, J. L. P.; Sobott, F.; Robinson, C. V. *Anal. Chem.* **2003**, *75*, 2208−2214.

stage pressure 4.0×10^{-3}—2.0×10^{-2} mbar, argon collision gas at a collision cell pressure of 2.5 bar. External calibration was achieved by using a 33 mg/mL aqueous solution of cesium iodide (Sigma, St. Louis, MO).

Mass Measurement and Assignment of Charge States. As a consequence of the large peak widths obtained from noncovalent assemblies, several charge state series can be fitted to the peak profile observed. To assign charge states unambiguously we carried out the following steps:

1. All charge states in the series were modeled as Lorentzian, Gaussian, or two-component Gaussian distributions using Sigmaplot 2001 (SPSS, Chicago, IL). The *m*/*z* values for all charge states were obtained from the centroid value of each curve.

2. A range of charge state assignments was iterated across the series.

3. The molecular weight of the complex was calculated from all measured *m*/*z* values using the charge state from each iterated assignment.

4. An average mass and standard deviation for the complex were calculated for all models.[7]

5. The lowest standard deviation model is identified from a plot of standard deviation versus charge state assigned; this identifies the model which has the lowest distribution of molecular masses around the mean.

6. Residuals, the differences between the observed and predicted responses, are elements of variation unexplained by the model. In this case, the residuals can be determined by comparing the calculated molecular weight from each *m*/*z* value in the assignment to the mean value.

Ribosome Preparation. Ribosomes were prepared using an established protocol,[17] adapted recently.[18] Aliquots for MS were buffer-exchanged using Bio-Rad micro biospin columns, MWCO 6 kDa, (Bio-Rad, Hercules, CA) into 1 M ammonium acetate solution and stored on ice.

Simulation of Spectra for Different Protein Compositions. Simulated spectra for known isotopic compositions were produced using Sigmaplot 2001 (SPSS, Chicago, IL). Gaussian curves were used to model individual peaks, with fwhm of 85 *m*/*z* units for the 30S. For the 70S, two components were modeled using fwhm of 207 and 110 *m*/*z*. These values are representative of the experimental data. The intensities of the peaks were based on the observed intensities of the peaks in the spectrum.

Results and Discussion

Observation of a Mass Increase in Large Protein Assemblies. To define the extent of adduct formation that is typical we considered a set of noncovalent assemblies reported previously from our laboratory, ranging from the ca. 365 kDa enolase assembly through to the ca. 800 kDa GroEL complex, Figure 2. Comparison of masses calculated from the known composition of complexes with masses calculated from peaks measured in the spectrum reveals that the measured masses are shifted to higher values. We noted that this mass increase is not related to the size of the complex, the largest mass shift being observed for the 390 kDa RNA polymerase complex and the smallest for the 800 kDa GroEL complex. Moreover the proteasome and GroEL are both open barrel structures of similar mass and solvent accessible surface areas (214 107 Å^2 and 297 094 Å^2, respectively).[19] After desolvation in the mass spectrometer, however, both assemblies retain differing amounts of adducts (0.3% and 1.0% mass increase observed for GroEL and proteasome, respectively). This is a rather surprising result, as assuming that the desolvation conditions in the instrument were optimum for each experiment, we would expect the physical characteristics of the protein complex, such as mass and surface area, to govern the extent of adduct formation and retention. However, the preservation of these assemblies relies upon the competing requirement to desolvate ions which has to be balanced by the need to retain noncovalent interactions in the gas phase. Consequently it is often impossible to desolvate complexes fully without disrupting their subunit interactions. In this case it is possible to remove more adducts from GroEL than the proteasome. Since the extent of adduct formation cannot be predicted from X-ray structures and given the differences in stability of protein complexes in the gas phase, we have elected to develop an approach to determine the extent of adducts that is independent of both the stability and extent of desolvation of the complex.

A Relationship between Distributions of Adducts and Increased Mass. We carried out a systematic investigation of a large subset of noncovalent complexes consisting of proteins of known sequence and with an established stoichiometry. Since a high level of accuracy is required in the mass measurement of these complexes only those spectra with well-defined Gaussian peak shapes were included in our dataset. Gaussian curves were fitted to the peaks in each spectrum as described in Materials and Methods, and centroid values were determined. After assignment of the charge, the overall mass of the complex was calculated from these centroid values. These values were compared to the molecular mass derived from the sequence in the protein database. As binding of small molecule adducts will both increase the mass and broaden the width of the charge states, we plotted the % mass increase against the average peak width (fwhm) for all charge states observed in the distributions, normalized to their *m*/*z* values, Figure 3A. Interestingly, we found that a linear relationship provides a good fit to the data with a high degree of confidence (correlation coefficient $R = 0.9556$).

It should be possible to test if the linear relationship is a good approximation of the relationship between peak broadening and mass increase by selecting an assembly of established protein composition and changing the experimental conditions to manipulate the extent of desolvation. To investigate this we used the well-studied *E. coli* $GroEL_{14-mer}$ under a series of different desolvation conditions.[15,25] Under the least efficient desolvation conditions we found the largest peak width and associated mass shift. As anticipated, with increasing collision energy more efficient desolvation was possible (see Supporting Information). With optimal desolvation conditions we found that it was possible to obtain a mass very close to that calculated from the sequence mass of the individual proteins that comprise the GroEL assembly. Most important, however, is the finding that the data points for these measurements, recorded under different desolvation conditions, fall well within the prediction interval

(17) Spedding, G. Oxford University Press: Oxford, 1990; pp 1—27.
(18) Clemons, W. M., Jr.; Brodersen, D. E.; McCutcheon, J. P.; May, J. L.; Carter, A. P.; Morgan-Warren, R. J.; Wimberly, B. T.; Ramakrishnan, V. *J. Mol. Biol.* **2001**, *310*, 827—843.
(19) Lee, B.; Richards, F. M. *J. Mol. Biol.* **1971**, *55*, 379—400.
(20) Ilag, L. L.; Videler, H.; McKay, A. R.; Sobott, F.; Fucini, P.; Nierhaus, K. H.; Robinson, C. V. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 8192—8197.
(21) Ilag, L. L.; Westblade, L. F.; Deshayes, C.; Kolb, A.; Busby, S. J.; Robinson, C. V. *Structure* **2004**, *12*, 269—275.
(22) Sharon, M.; Witt, S.; Felderer, K.; Rockel, B.; Baumeister, W.; Robinson, C. V. *J. Biol. Chem.* **2006**, *281*, 9569—9575.
(23) Guex, N.; Peitsch, M. C. *Electrophoresis* **1997**, *18*, 2714—2723.
(24) Schurig, H.; Rutkat, K.; Rachel, R.; Jaenicke, R. *Protein Sci.* **1995**, *4*, 228—236.
(25) van Duijn, E.; Bakkes, P. J.; Heeren, R. M. A.; van den Heuvell, R. H. H.; van Heerikhuizen, H.; van der Vies, S. M.; Heck, A. J. R. *Nat. Methods* **2005**, *2*, 371—376.

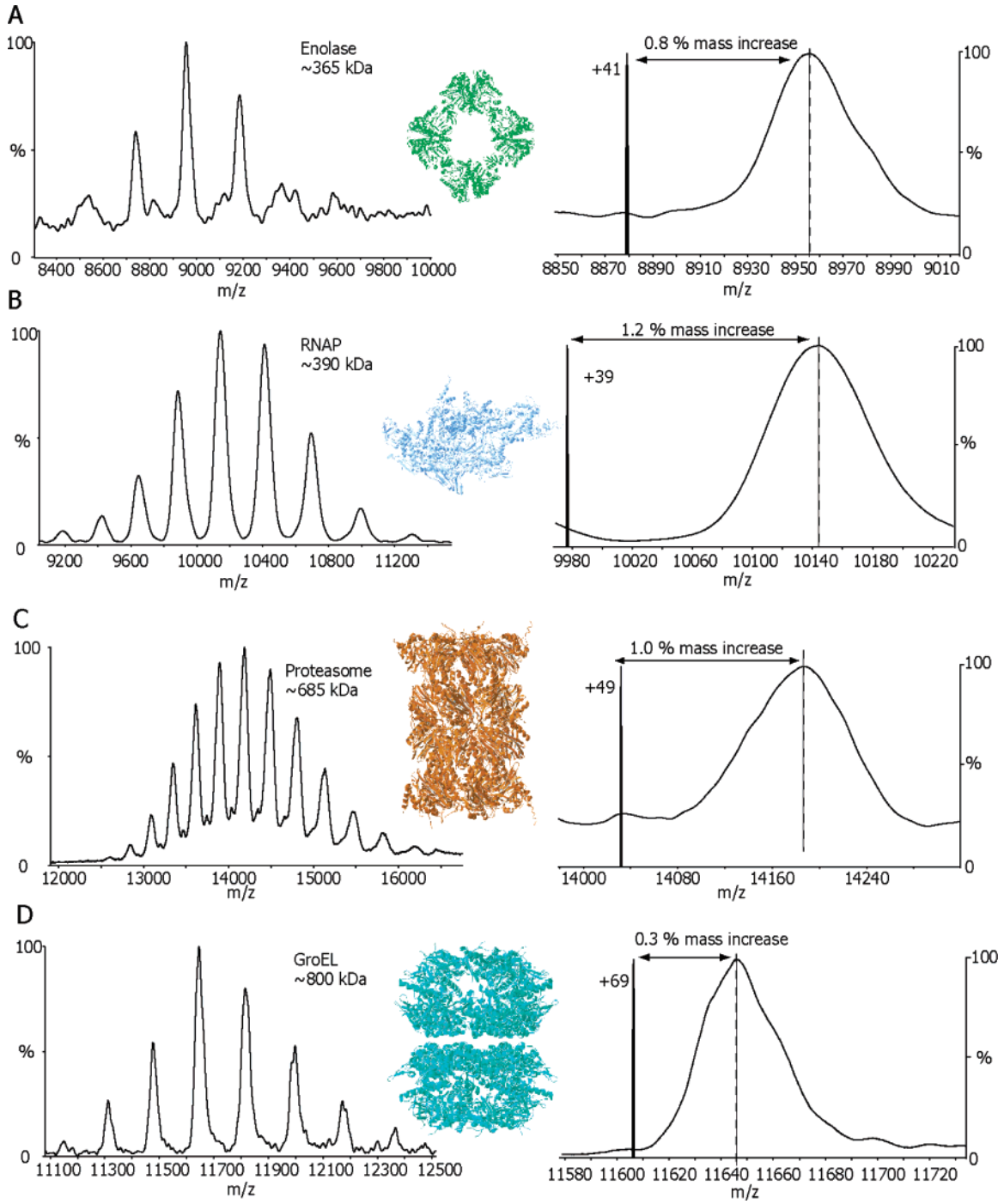

Figure 2. (A−D) Mass spectra of four multiprotein assemblies, octameric enolase,[20] RNAP,[21] 20S proteasome minus an α-subunit,[22] and GroEL.[15] An expansion of the base peak including the isotope model calculated from the sequence mass reveals that each charge state has an associated mass shift that is independent of the mass of the complex. Crystal structures that are drawn to scale have been included for the four complexes presented. (Protein Data Bank coordinates, 1GRL, 1E9I, 1IRV, and images were prepared using the Swiss Pdb Viewer.[23]) Enolase exists as an octamer in thermophillic bacteria; here the dimer from *E. coli* was oriented to give an octameric structure consistent with electron microscopy images.[24]

derived for the linear relationship shown in Figure 3A. Repeating the regression analysis to include three charge states from the $GroEL_{14-mer}$ (+68, +69, and +72) under different desolvation conditions yielded a linear relationship closely similar to the one deduced above with all points within the 99% prediction interval (correlation coefficient, $R = 0.9507$), Figure 3B. All three charge states examined obey the same relationship indicating that the method is charge state independent. These results recorded for the $GroEL_{14-mer}$ serve to validate our model by demonstrating that the extent of desolvation is the primary cause of the increased mass and that the correction we apply, derived solely from the average peak widths of the charge states, enables a reliable determination of the mass of the complex with associated adducts.

It is important to emphasize that the *relationship* between peak width and adduct retention is not specific to a particular set of experimental parameters or a particular mass spectrometer. The underlying physical basis of the relationship shown in Figure 3 is two-fold. First, the probability of observing protein complexes containing a large distribution of adduct ions,

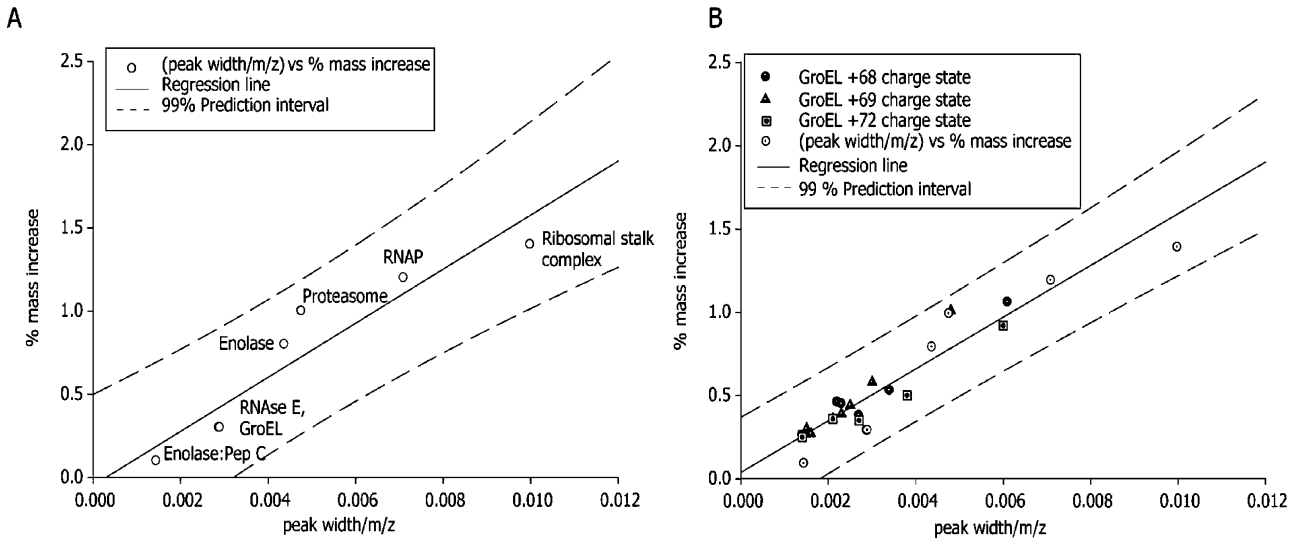

Figure 3. (A) Plot of % mass increase versus peak width (normalized to *m*/z) for a series of seven complexes, enolase:pep C,[26] the catalytic domain of RNAse E,[27] $GroEL_{14mer}$,[15] enolase,[20] RNAP,[21] the ribosomal stalk complex,[20] and the 20S proteasome minus α-subunit.[22] The dashed line indicates the 99% prediction interval for the regression line (line equation from regression analysis: % mass increase = 160(peak width/*m*/z) − 0.03). (B) Plot of % mass increase versus peak width (normalized to *m*/z) for the seven complexes in plot A plus three representative charge states of GroEL obtained under different desolvation conditions (line equation from regression analysis: % mass increase = 156(peak width/*m*/z) + 0.03).

evidenced by a large peak width in the mass spectrum, is increased as the number of adducts bound to the protein complex increases. Second, the average mass of typical adducts that are common to protein complexes of biological origin is largely invariant. Based on the above, the gradient of the relationship shown in Figure 3 can be correlated to the average mass of adducts typically associated with protein complexes, and the *x*-intercept approximates to the theoretical line-width of the charge state, derived from the isotopic composition of the protein complex. As a result, no knowledge of the physical attributes of the protein complex that may affect adduct removal in the desolvation process is required in order to apply a correction factor derived from Figure 3. Consequently, it is possible to estimate the mass increase of an unknown, based only on the width of the peaks observed in the spectrum.

Assigning Charge States to Mass Spectra of the 30S Ribosomal Subunits. Having deduced this relationship between peak width and adducts we applied this approach to one of the most challenging systems that we have encountered, the intact 30S ribosomal subunit from *Thermus thermophilus*. The 30S ribosomal subunit has an approximate mass of 800 kDa and consists of up to 20 proteins and 1 RNA molecule (the 16S rRNA). Although high-resolution X-ray analyses have revealed much of the static composition of the ribosome,[28−31] the exact enumeration of ribosomal proteins has met with some difficulties.[32] Careful control of the conditions in the mass spectrometer

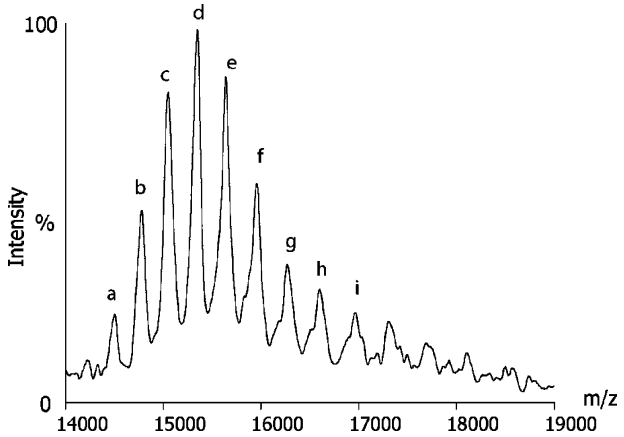

Figure 4. Electrospray mass spectrum of the small ribosomal subunit from *Thermus thermophilus*. The base peak in the charge state series is labeled (d). Charge states labeled (a−e) represent a well-defined peak shape consistent with one protein composition. Heterogeneity is evident for charge states g−i.

leads to remarkably well resolved charge states (a−i) (Figure 4). Using the procedures outlined above Gaussian curves were fitted to determine the apex of these peaks and *m*/z values were estimated for nine of the peaks (a−i). Using these *m*/z values we determined the mean molecular mass and standard deviation for a series of 20 different charge state assignments. To determine the charge state assignment which has the lowest standard deviation we employed an approach reported previously and plotted the standard deviation of the calculated mass against the charge state assignment[7] (Figure 5A). The graph indicates that the model that assigns +49 charge state to the base peak (peak d) in the distribution has the lowest standard deviation when compared to the 19 other possible assignments. The U-shaped profile of this plot, however, together with the magnitude of the standard deviation values associated with all 20 assignments implies that none of the models adequately describe the data. To investigate this further we plotted the residuals calculated from the difference between the mass

(26) Callaghan, A. J.; Aurikko, J. P.; Ilag, L. L.; Gunter Grossmann, J.; Chandran, V.; Kuhnel, K.; Poljak, L.; Carpousis, A. J.; Robinson, C. V.; Symmons, M. F.; Luisi, B. F. *J. Mol. Biol.* **2004**, *340*, 965−979.
(27) Callaghan, A. J.; Grossmann, J. G.; Redko, Y. U.; Ilag, L. L.; Moncrieffe, M. C.; Symmons, M. F.; Robinson, C. V.; McDowall, K. J.; Luisi, B. F. *Biochemistry* **2003**, *42*, 13848−13855.
(28) Harms, J.; Schluenzen, F.; Zarivach, R.; Bashan, A.; Gat, S.; Agmon, I.; Bartels, H.; Franceschi, F.; Yonath, A. *Cell* **2001**, *107*, 679−688.
(29) Schluenzen, F.; Tocilj, A.; Zarivach, R.; Harms, J.; Gluehmann, M.; Janell, D.; Bashan, A.; Bartels, H.; Agmon, I.; Franceschi, F.; Yonath, A. *Cell* **2000**, *102*, 615−623.
(30) Ban, N.; Nissen, P.; Hansen, J.; Moore, P. B.; Steitz, T. A. *Science* **2000**, *289*, 905−920.
(31) Wimberly, B. T.; Brodersen, D. E.; Clemmons, W. M., Jr; Morgan-Warren, R. J.; Carter, A. P.; Vonrhein, C.; Hartsch, T.; Ramamkrishnan, V. *Nature* **2000**, *407*, 327−339.
(32) Liljas, A. World Scientific Publishing Co. Pte. Ltd.,: 2004; pp 42−46.

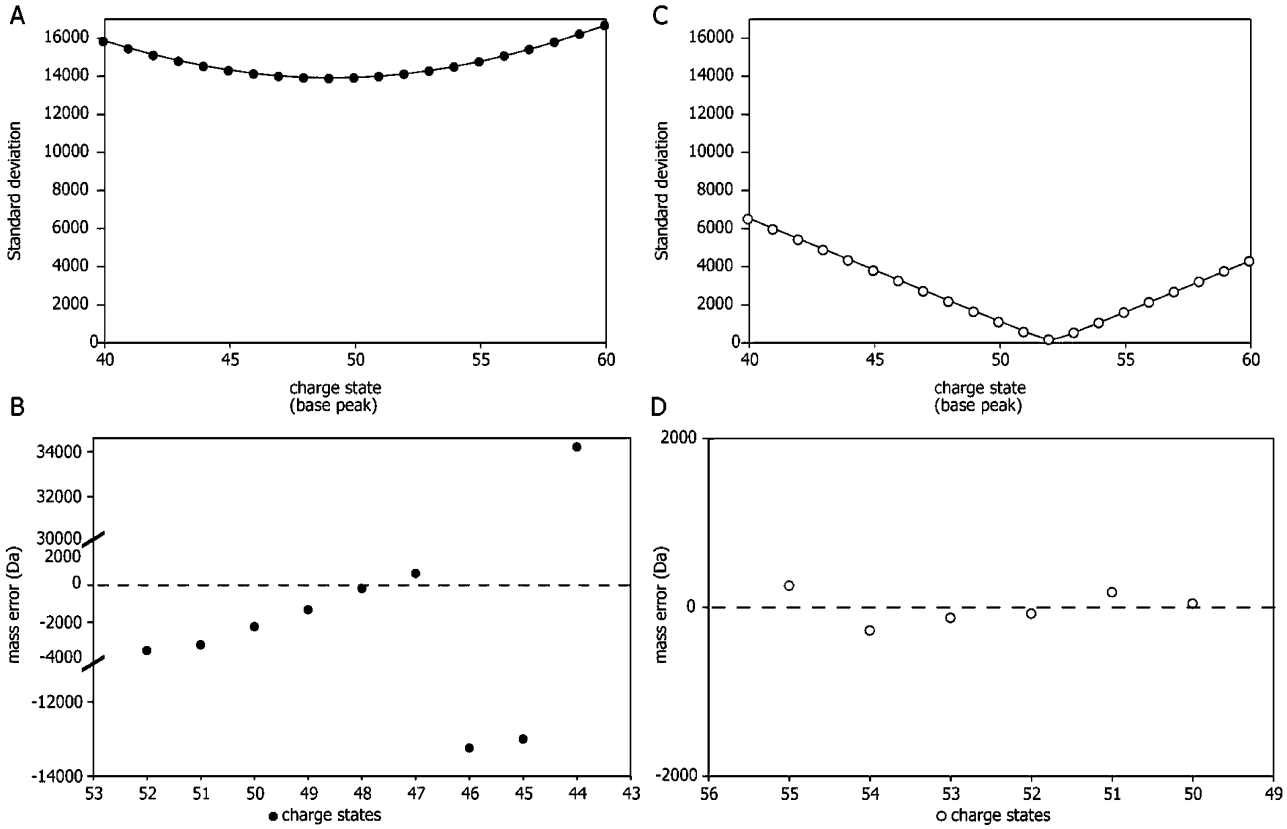

Figure 5. Statistical analysis of various charge state assignments. (A) Plot of standard deviation versus the charge state assignment of the base peak (d) for 20 charge state distributions. Although the standard deviations of all iterations are high, the lowest value is the model that assigns the +49 charge state to peak d. (B) Plot of residuals versus charge state for the model that assigns the +49 charge state to peak d. It is clear that there are large errors in the three peaks at highest *m*/*z*; they are an order of magnitude larger than the other six charge states in this model. (C) The three peaks g−i have been excluded, and the analysis was performed as described above with this reduced data set. The standard deviation values are much reduced for all iterations; the +52 model has the lowest standard deviation. (D) Plot of residuals versus charge state for the model that assigns the +52 charge state to peak d. For this model residual errors are small and randomly distributed for the first six charge states. This suggests that the model that assigns a +52 charge to peak d adequately describes the data.

obtained for each charge state and the mean value found for the lowest standard deviation model in which peak d is +49. (Figure 5B)

If the charge state assignment of peak d as +49 is an adequate fit for all the peaks in the spectrum we would expect it to err by predicting values higher than and lower than the mean value from the lowest standard deviation model with equal probability. In addition, the magnitude of the error should not trend with the charge state. Deviations from this behavior therefore imply inadequacies in the model. A plot of the residuals calculated from our "best fit" assignment indicates that the error is charge state dependent and gives rise to one very large positive value (+44) and two large negative values (+45 and +46). These three peaks have residual errors that are an order of magnitude greater than the six other peaks calculated from the model where peak d is assigned as +49. We therefore disregard these three values (peaks g−i) and repeated the iteration, as described above. If we plot the revised standard deviation, not including these three charge states, we obtain a V-shaped profile. This allows us to define with confidence the best fit of the charge assignment (Figure 5C). A plot of the residuals for this model errs in a random fashion, and the magnitude of the errors is much reduced compared with the model before rejection of the peaks g−i. The standard deviation is also considerably reduced. We, therefore, assign the base peak in the spectrum (d) as the +52 charge state of the 30S subunit.

Accurate Mass Determination of the 30S Ribosomal Subunit. Using the "best fit" assignment of the base peak (d) as +52 we obtain a measured mass of 797 605 ± 119 Da. The average peak width in the mass spectrum has a fwhm value of 85 *m*/*z*. Using this value and the relationship between peak width and mass increase deduced in Figure 3 we estimate a mass increase of 0.9% due to the formation of adducts. Calculation of the anticipated mass of the 30S subunit using the established protein and RNA sequences gives a mass of 818 784 Da. The mass of the ribosomal subunit calculated from sequences and adjusted for adducts is therefore 826 153 Da. This is larger than the measured value by 28548 Da and the difference does not correspond to any 30S ribosomal protein. Biochemical analyses suggest that preparations of *E. coli* ribosomes can contain less than stoichiometric quantities of protein S1.[33] In addition we have shown previously that excitation of the 30S subunit in the mass spectrometer releases protein S6.[20] This could in principle give rise to four different charge state series with (i) the full complement of proteins, (ii) the 30S subunit−S1, (iii) 30S subunit−S6, and the (iv) 30S subunit−S1 and −S6 (Figure 6). Interestingly the mass determined for the peaks a−f is consistent with an average of the masses of the intact 30S and the 30S−S1, implying that both species are present but not resolved within the spectrum (Table 1).

(33) Subramanian, A. R. *J. Mol. Gen. Genet.* **1977**, *158*, 1−9.

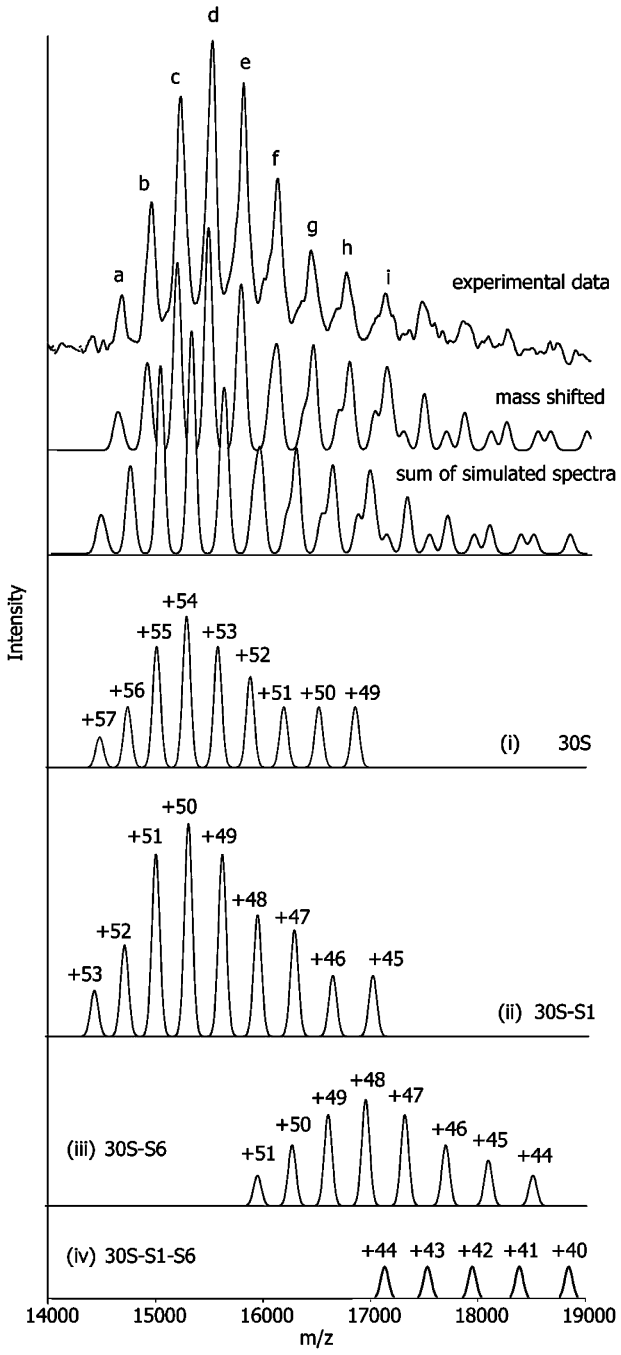

Figure 6. Simulations of the charge state distributions for the four possible compositions (i) the 30S subunit, (ii) the 30S subunit−S1, (iii) the 30S subunit−S6, and (iv) the 30S subunit−S1 and −S6. Simulations were produced using the sequence masses for the known components of the complex, and intensities were adjusted to fit the experimental data. The simulations were summed, and the mass shifted according to the relationship derived in Figure 3 such that they could be compared directly with the experimental data.

If we now consider the charge states at higher *m*/*z* (g−i), which were rejected from our analysis initially since they did not appear to belong to charge state series a−f, we find that rather than a near Gaussian distribution of charges states anticipated for a single component[34] the spectrum is asymmetric with an extended series of low intensity charge states at high *m*/*z*. Moreover shoulders are evident for charge states g−i implying overlapping charge state series similar to those simulated in Figure 1B above. If the charge state series is composed of overlapping series, we would anticipate that it would be most evident for the lower charge states (g−i) where the *m*/*z* differences are largest. To investigate the contribution to the mass spectrum of ribosomes with different protein components we have simulated the spectra for the four different compositions (i−iv) detailed above.

Modeling the Mass Spectrum of the 30S Ribosomal Subunit. Simulation of the individual charge-state distributions for the two 30S subunits (S1 bound and S1 free) together with the related versions, in which both have lost protein S6, are shown, Figure 6. The individual spectra were summed, and the intensities were adjusted until the best fit between experimental and modeled mass spectra were obtained. The mass correction deduced from Figure 3 was then applied to the simulation so that it could be compared directly with our experimental spectrum.

The results show that there is good agreement between the simulated and experimental data for the peaks a−e and that the shoulders observed in the experimental data for peaks g−i are reproduced in the simulation. The relative intensities of the various simulations allow us to estimate the percentage of subunits in which S1 is absent. We conclude that approximately 50% of the 30S subunits contain the ribosomal protein S1. Although we anticipated that the intensities of 30S subunits from which S6 has dissociated would reflect the intensities of the intact 30S and 30S−S1 populations from which they are derived, this is not what we observed. However given that the base peak (d) is assigned as +54 and +50 for the 30S and the 30S−S1, respectively we attribute the discrepancy in the dissociation products, formed by loss of S6, to the mean charge state difference between the two precursors. This is evident from the lower intensity of the simulated spectra of the 30S−S1−S6 compared to that of the 30S−S6 spectrum. These simulated spectra illustrate how the four distinct populations combine to give the spectrum recorded for the 30S assembly. Moreover, by adjusting the relative intensities we are able to estimate the contribution of each of the different populations in the ribosome preparation.

Tandem Mass Spectrometry Confirms the Presence of Different Populations. To confirm the presence of overlapping populations of 30S ribosomes we carried out tandem MS. Isolation of a region of the spectrum corresponding to the intact 30S subunit and activation in the collision cell at 80 V increases the degree to which the ions are desolvated. Under these conditions, the presence of overlapping charge states is revealed within the unresolved peaks in the mass spectrum, Figure 7. Comparison with the appropriate charges states from the simulated spectra of the 30S and 30S−S1 ribosomes shows that they are in good agreement with the experimental data. This provides additional evidence for the origin of the overlapping 30S and 30S−S1 peaks (a−e) in the mass spectrum, revealed by mass measurement. If we now increase the collision cell voltage from 80 V to 200 V we induce dissociation of the 30S subunit. The low molecular mass dissociation product is found to correspond to the protein S6, as observed in our previous studies.[25] However under these conditions we are also able to observe the "stripped" complex corresponding to the 30S assembly from which S6 has dissociated. Mass measurement

(34) Dobo, A.; Kaltashov, I. A. *Anal. Chem.* **2001**, *73*, 4763−4773.

Table 1. Masses Calculated from Known Sequence Data Are Adjusted for Adduct Binding and Compared with Those Measured in Mass Spectra

subunit composition	sequence mass (Da)	sequence mass after adding adjustment (Da)	measured mass (Da)	Δmass (Da) (measured – adjusted mass)
intact 30S	818 784	826 153		−28 548
30S−S1	758 814	765 643		31 962
30S−S6	806 943	814 205	797 605 ± 119	−16 600
30S−S1−S6	747 062	753 786		43 819
average of 30S and 30S−S1	788 799	795 898		1707
70S	2 295 888	2 323 439	**2 321 655 ± 1562**[a]	−1784
70S−S1	2 235 918	2 251 569	**2 256 508 ± 5334**[a]	4939
30S−S6 (MSMS product)	806 943	810 171	810 155 ± 98	−16

[a] For 70S ribosomes it is possible to obtain two measurements of the mass from the *m*/*z* values determined by the two-component Gaussian regression. The fact that these two measured masses (bold) are closely similar to those calculated from the sequence masses after adjustment for adducts increases our confidence in our model for the mass spectrum of 70S ribosomes.

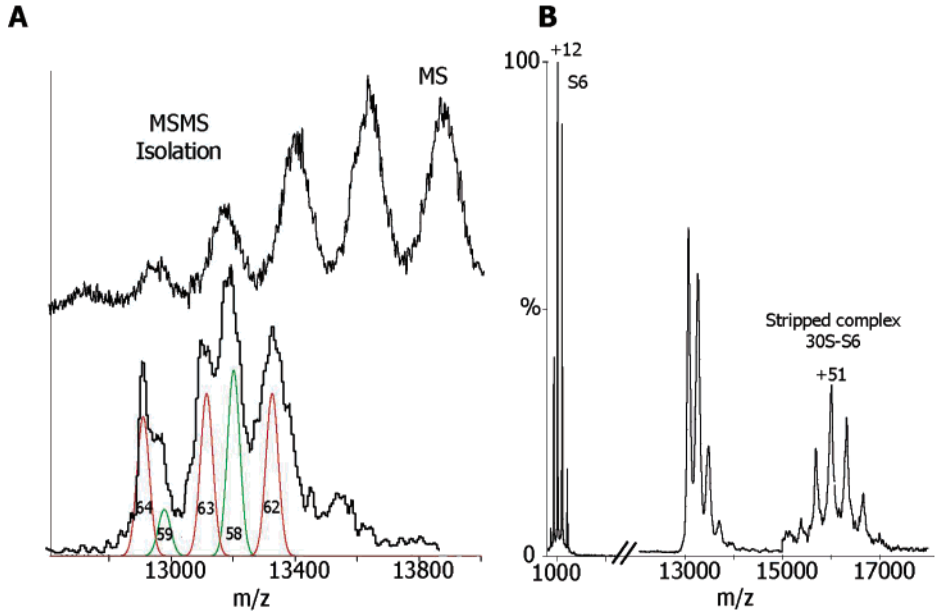

Figure 7. (A) Expansion of the region in the mass spectrum assigned to 30S subunits and acceleration through the collision cell at 80 V. The increased acceleration improves the desolvation of the complex revealing the presence of at least two species. Simulated spectra for the 30S subunit (charge states +62 to +64 red) and the 30S−S1 (+58 and +59 green) were created using masses that include the estimated mass of adducts. These simulated charge states coincide with the two populations observed experimentally. (B) Tandem mass spectra were recorded by isolating the highest charge states of the intact 30S subunits and raising the collision cell voltage to 200 V. Only one protein is released (S6). The "stripped" complex at *m*/*z* ∼16 000 corresponds in mass to 30S−S6. The intensity of the stripped complex has been magnified 6-fold.

of these peaks, using the procedure established here, confirms that they arise from the 30S−S6. No contribution from peaks assigned to loss of both S1 and S6 could be detected. We attribute this observation to the higher charge states of the 30S (62/63/64+) compared with the 30S−S1 (57/58+) and the predominance of the intact 30S charge states in the tandem MS isolation, Figure 7. In summary, the increased desolvation corroborates the presence of populations of ribosomes, with and without S1. The tandem mass spectrometry demonstrates facile loss of S6 to yield a population of 30S−S6 confirmed by mass measurement of the stripped complex (Table 1). Taken together these tandem mass spectrometry experiments support our conclusions derived from mass measurement of the intact protein complex.

Accurate Mass Measurement of the Intact 70S Ribosome. The insight that we have gained from mass measurement of 30S subunits prompted us to study the intact 70S ribosomal particles from *Thermus thermophilus*. These intact particles are close to being 3 times larger than 30S subunits with an approximate mass of 2.3 MDa and consist of up to 58 proteins and 3 RNA molecules (5S, 16S, and 23S ribosomal RNA). With careful control of the instrument conditions, well resolved peaks can be obtained (a−i) (Figure 8).

Using the mass measurement approach described for the 30S subunit, *m*/*z* values were estimated for all nine peaks (a−i) and the "best fit" model assigns the base peak e as +90. From this assignment we calculate the molecular mass 2 325 148 ± 586 Da. The base peak in the mass spectrum has a peak width of 294 *m*/*z*, and from Figure 3 we would anticipate an increase of 1.8% on the mass calculated from protein and RNA sequences to give a mass of 2 337 213 Da. This mass however is 12 065 Da greater than the 2 325 148 Da measured from the spectrum. One possible explanation for the deviation from our measure-

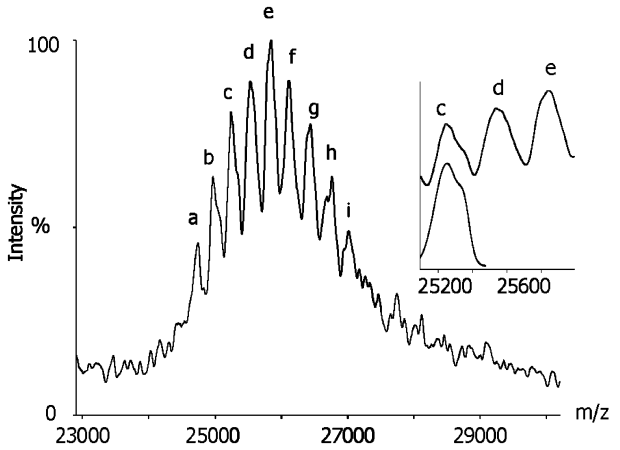

Figure 8. Mass spectrum of 70S ribosomes from *Thermus thermophilus*. The base peak in the charge state series is labeled (e). An expansion of the *m*/*z* region containing peaks c, d, and e is shown in the inset with the two-component Gaussian fitting of peak c.

ment is that there is more than one population of 70S particles, as deduced for 30S subunits.

Heterogeneity in 70S Ribosomes. In addition to the discrepancy in the mass, the exceptionally broad peak width of 70S charge states also suggests that more than one species is contributing to the mass spectrum. Our approach is limited to species that can be adequately resolved in the mass spectrum. Shoulders are evident on peaks b, c, and h. If we carry out two-component Gaussian fitting on these peaks (an example of which is shown for peak c, Figure 8 inset), it is possible to resolve two components. An average of the peak width estimated from the three two-component Gaussian fits yields values of 207 *m*/*z* and 110 *m*/*z* for the fwhm of the two components. For these peak widths and centroid values, mass increases of 1.2% and 0.7%, respectively, are predicted from Figure 3. Applying two separate adjustments to the sequence masses calculated for the intact 70S and 70S−S1 ribosomes gives rise to two populations with molecular masses of 2 323 439 and 2 251 569 Da. The difference between these two values (71 780 Da) could be explained by any number of combinations of small ribosomal proteins. However given that the peaks are resolved in the spectra, indicating that there are not multiple combinations of different substoichiometric ribosomes, it is likely that this mass difference corresponds to incomplete binding of the individual ribosomal protein S1 (59 970 Da) shown to be substoichiometric in 30S subunits in this study. We also considered partial loss of the 96 kDa stalk complex, as previous studies have shown this to be particularly labile in the gas phase.[20]

We simulated spectra of 70S particles with and without the stalk complex together with a number of alternative models with differing degrees of adduct formation on the ribosome populations. None of these simulations provide an adequate fit to the experimental data (see Supporting Information). However given that we have already established heterogeneity in the 30S subunit and ascribed this to the presence and absence of S1 and since the 30S and 50S together constitute 70S ribosomes, it is reasonable to assume that the intact particle must also contain variant 30S subunits. We would anticipate these two populations would not be resolved within the context of the intact particle, particularly since considerable overlap was observed in the spectrum of the 30S subunit alone. The most likely explanation

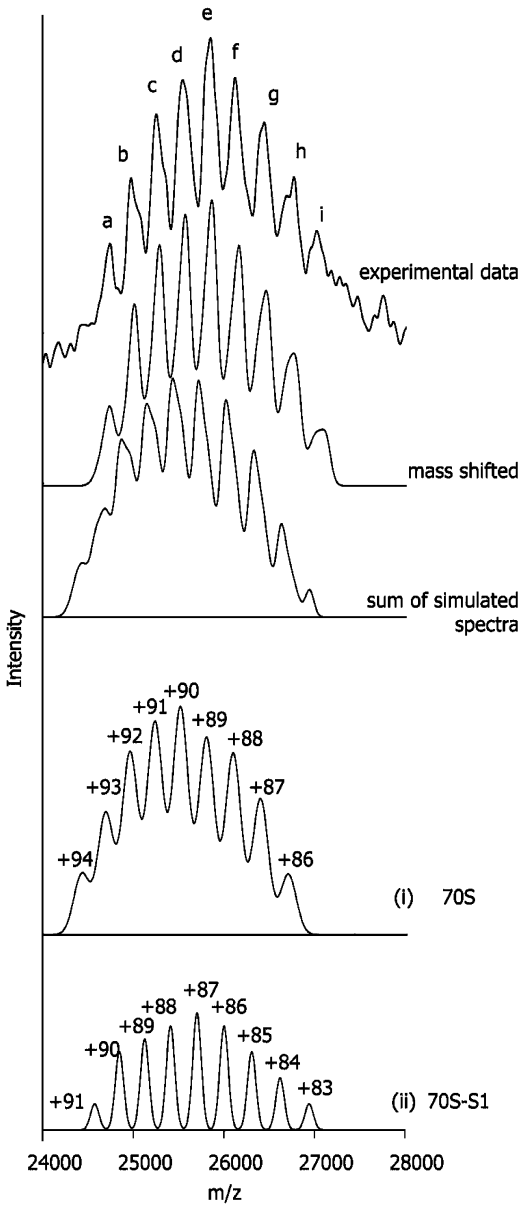

Figure 9. Simulations of the charge state distributions for two possible compositions of 70S ribosomes: (i) 70S ribosome with the full protein complement; (ii) the 70S ribosome−S1. Simulations were produced using the sequence masses for the known components of the complex, and intensities were adjusted to fit the experimental data. The simulations were summed and mass shifted such that they could be compared directly with the experimental data. It is important to note that the resolution of the simulated spectra is increased upon application of the mass shift. This arises since two different adjustments are applied to the 70S and 70−S1 (1.2% and 0.7%, respectively).

for the two masses determined for 70S particles is the presence and absence of S1, as deduced for 30S subunits.

Modeling the Mass Spectrum of the Intact 70S Ribosome. To explore how the 70S and 70S−S1 ribosomes could combine to produce the mass spectrum observed, we simulated the individual charge-state distributions for these species (Figure 9). The intensity of the peaks in the different simulations was adjusted to fit the experimental data and summed. The adduct adjustment was then applied to the simulation so that it could be compared directly with our experimental spectrum. The results show that there is surprisingly good agreement between the simulated and experimental data, given the size of this

particle. Although the shoulders observed in the experimental data for peaks b, c, and h are not evident in the simulation, the broadening resulting from the overlapping signals is clearly reproduced. Moreover, given that the calculated errors between the adduct adjusted and measured masses of the 70S and 70S−S1 ribosomes are remarkably small, differing by less than 2 and 5 kDa, respectively (<0.1 and 0.3%; see Table 1), we conclude that this simple model is the best fit to our data. Therefore, the major contribution to heterogeneity in 70S ribosomes can be adequately explained by the presence of two populations, with and without S1, as revealed for the 30S subunit.

Conclusions

In this study we have examined a range of noncovalent complexes and shown that there is no obvious correlation between the factors affecting adduct formation in noncovalent assemblies and their physical properties including mass or solvent accessible surface area. From measurement of the peak width, however, we show that it is possible to estimate the mass increase of an unknown, based solely on the width of the charge states observed in the spectrum. Applying this "adduct adjustment" approach we were able to determine the protein composition of 30S ribosomal subunits and intact 70S ribosomes. The presence of protein S1 in *Thermus thermophilus* was, until recently, not established,[35] and its existence has been called into question.[36] We have confirmed the presence of S1 on 30S subunits as well as two further species corresponding to the dissociation of protein S6 from the S1-bound and S1-free 30S ribosomal species present in solution. We have also defined the protein composition of intact 70S ribosomes demonstrating that the majority of the particles have a full protein complement, including the flexible stalk complex for which very little data exists.

It has been established for some time that applying standard software approaches to determining the molecular mass of large assemblies is not effective. Commercial software packages such as the deconvolution program based on maximum entropy[37] are unable to provide the correct multiple component solution to the mass spectra of the 30S subunits. Moreover from our initial examination of the mass spectra of this particle, and from procedures that we had applied previously,[7] it was not possible to obtain either the molecular mass or stoichiometry of the protein composition without ambiguity. By applying the procedures outlined here we have delineated the various contributions to the overall mass spectrum and obtained reliable molecular masses for all the constituent assemblies.

To date 30S subunits and intact 70S ribosomes are two of the most heterogeneous assemblies for which accurate molecular masses have been determined. It is clear from this study that 2.3 MDa 70S ribosomes are currently at the upper limit for high-accuracy molecular mass determination, particularly given the potential for heterogeneity in 3 large RNA molecules and over 50 different proteins. The 70S data clearly illustrate that the resolution of the charge states ultimately limits our ability to identify the ionized components of protein complexes. It is important to note however that this limitation is not imposed by the mass analyzer employed but rather by the electrospray process itself which provides both highly charged and poorly desolvated ions in a manner that is dependent upon the protein complex under investigation.

The absence of effective methods for determining the mass of noncovalent protein assemblies has hindered progress of this emerging field. Given that recording mass spectra of such large and heterogeneous particles requires considerable investment in optimization of solution and mass spectrometry conditions it is imperative that we extract the maximum information content from the resulting mass spectra. We anticipate therefore that the data analysis approach described here, which involves combining adjustment for adducts, accurate mass measurement, simulation of spectra, and tandem MS, will advance significantly the mass spectrometry of large heterogeneous macromolecular assemblies.

Acknowledgment. The authors thank Venki Ramakrishnan and Ann Kelley (Medical Research Council, Cambridge, UK) for kind gifts of ribosomes and Helena Hernández for critical review. A.R.M. acknowledges the Medical Research Council and Syngenta, C.V.R., the Walters Kundert Charitable Trust, and B.T.R., the Research Councils Basic Technologies Program.

Supporting Information Available: Stack Plot of GroEL spectra obtained under varying desolvation conditions, alternative models for 70S ribosomes, and spectra adjusted for differing degrees of adduct formation. This material is available free of charge via the Internet at http://pubs.acs.org.

JA061468Q

(35) Shiryaev, V. M.; Selivanova, O. M.; Hartsch, T.; Nazimov, I. V.; Spirin, A. S. *FEBS Lett.* **2002**, *525*, 88.
(36) Tsiboli, P.; Herfurth, E.; Choli, T. *Eur. J. Biochem.* **1994**, *226*, 169−177.
(37) Ferrige, A. G.; Seddon, M. J.; Green, B. N.; Jarvis, S. A.; Skilling, J. *Rapid Commun. Mass Spectrom.* **1992**, *6*, 707−711.