

Accelerated Articles

Quantitative ESI-TOF Analysis of Macromolecular Assembly Kinetics

Anne E. Bunner,[†] Sunia A. Trauger,[‡] Gary Siuzdak,[‡] and James R. Williamson^{*,†}

Departments of Molecular Biology and Chemistry, The Skaggs Institute for Chemical Biology, and Scripps Center for Mass Spectrometry, The Scripps Research Institute, 10550 North Torrey Pines Rd, La Jolla, California 92037

The *Escherichia coli* small (30S) ribosomal subunit is a particularly well-characterized model system for studying in vitro self-assembly. A previously developed pulse-chase monitored by quantitative mass spectrometry (PC/QMS) approach to measuring kinetics of in vitro 30S assembly suffered from poor signal-to-noise and was unable to observe some ribosomal proteins. We have developed an improved LC–MS based method using quantitative ESI-TOF analysis of isotope-labeled tryptic peptides. Binding rates for 18 of the 20 ribosomal proteins are reported, and exchange of proteins S2 and S21 between bound and unbound states prevented measurement of their binding kinetics. Multiphasic kinetics of 3' domain proteins S7 and S9 are reported, which support an assembly mechanism that utilizes multiple parallel pathways. This quantitative ESI-TOF approach should be widely applicable to study the assembly of other macromolecular complexes and to quantitative proteomics experiments in general.

Self-assembling ribonucleoprotein complexes (RNPs) are involved in many essential cellular processes such as transcription, mRNA splicing, post-transcriptional regulation, and translation. The *Escherichia coli* small (30S) ribosomal subunit is a particularly well-characterized RNP model system for studying self-assembly.¹ The 30S subunit, which is made up of 20 ribosomal proteins (r-proteins) and a single 16S rRNA (rRNA),² self-assembles in vitro³ through a series of hierarchical RNA folding and protein binding

events.⁴ The Nomura assembly map^{4,5} (Figure 1A) describes protein binding under equilibrium conditions, based on experiments using different combinations of ribosomal proteins in partial reconstitutions. The order of protein binding suggested by these thermodynamic dependencies is consistent with kinetic data since primary binding proteins, those that bind directly to the RNA, tend to bind faster than secondary and tertiary proteins, which depend on other proteins in order to bind at equilibrium.⁶ The 30S subunit is organized into three domains that can be reconstituted independently: the 5' domain, the central domain, and the 3' domain.^{7–9} In vitro assembly is faster closer to the 5' end of the RNA,⁶ paralleling the cotranscriptional protein binding that likely occurs in vivo.¹⁰ The mechanism of in vitro assembly involves multiple parallel pathways rather than a single defined order of protein binding.^{11,12} Additional kinetic data will provide further insights into the mechanism of the in vitro 30S ribosomal subunit assembly.

Traditional methods of monitoring in vitro complex formation often require the use of fluorescent or radioactive tags that can be difficult to attach and can necessitate additional control experiments. Mass spectrometry can directly detect most biological molecules without the use of such tags. Mass spectrometry of stable-isotope labeled proteins and peptides has become a widespread technique for measuring relative protein levels be-

* To whom correspondence should be addressed.

[†] Departments of Molecular Biology and Chemistry and the Skaggs Institute for Chemical Biology.

[‡] Scripps Center for Mass Spectrometry.

(1) Kaczanowska, M.; Rydén-Aulin, M. *Microbiol. Mol. Biol. Rev.* **2007**, *71*, 477–494.

(2) Hardy, S. J. S.; Kurland, C. G.; Voynow, P.; Mora, G. *Biochemistry* **1969**, *8*, 2897–2905.

(3) Traub, P.; Nomura, M. *Proc. Natl. Acad. Sci. U.S.A.* **1968**, *59*, 777–784.

(4) Held, W. A.; Ballou, B.; Mizushima, S.; Nomura, M. *J. Biol. Chem.* **1974**, *249*, 3103–3111.

(5) Culver, G. M. *Biopolymers* **2003**, *68*, 234–249.

(6) Powers, T.; Daubresse, G.; Noller, H. F. *J. Mol. Biol.* **1993**, *232*, 362–374.

(7) Weitzmann, C. J.; Cunningham, P. R.; Nurse, K.; Ofengand, J. *FASEB J.* **1993**, *7*, 177–180.

(8) Recht, M. I.; Williamson, J. R. *J. Mol. Biol.* **2004**, *344*, 395–407.

(9) Samaha, R. R.; O'Brien, B.; O'Brien, T. W.; Noller, H. F. *Proc. Natl. Acad. Sci. U.S.A.* **1994**, *91*, 7884–7888.

(10) de Narvaez, C. C.; Schaup, H. W. *J. Mol. Biol.* **1979**, *134*, 1–22.

(11) Talkington, M. W.; Siuzdak, G.; Williamson, J. R. *Nature* **2005**, *438*, 628–632.

(12) Nomura, M. *Science* **1973**, *179*, 864–873.

tween samples.^{13,14} Two main challenges in mass spectrometry of stable-isotope labeled proteins and peptides are quantitation and identification. One of the most common methods for quantitating pairs of labeled and unlabeled isotopomers in liquid chromatography coupled mass spectrometry (LC–MS) data uses extracted ion chromatograms (EICs) of the labeled and unlabeled monoisotopic peaks.^{15,16} Although this method accounts for the whole width of the monoisotopic peak in the time dimension, it does not account for the intensities of heavier isotopomers in the isotope distribution of either species. Tandem mass spectrometry (MS/MS) is a powerful and widely used method for identifying peptides¹⁷ using accurate mass and time tags.^{18,19} However, this approach is not optimal for quantitative analysis of stable-isotope labeled samples because the same MS/MS data used for identification cannot easily be used for relative quantitation. This approach also limits the number of peptides that can be identified in routine samples to the set that have already been observed in MS/MS experiments.

Previous work in our laboratory used isotopically labeled r-proteins to perform quantitative in vitro studies of 30S ribosomal subunit assembly using MALDI of intact proteins.¹¹ These pulse-chase experiments monitored by quantitative mass spectrometry (PC/QMS) were initiated by mixing ¹⁵N labeled r-proteins with 16S rRNA (Figure 1B). The assembly reaction was then chased with excess unlabeled proteins at various time points, and 30S formation was allowed to proceed to completion. In the resultant 30S subunits, the fraction of ¹⁵N protein was equal to the fraction of that protein bound to the rRNA before the chase was added. The relative amounts of ¹⁴N and ¹⁵N protein were then quantitated using MALDI-MS of the protein mixtures extracted from the 30S subunits. The r-protein binding rates obtained using MALDI-PC/QMS analysis of intact proteins¹¹ (Figure 1A) were consistent with previous qualitative kinetic data based on chemical probing.⁶

The MALDI data from intact r-proteins suffered from poor signal-to-noise, especially for the larger proteins, and proteins S2 and S7 were not visible at all in the MALDI spectra. As an alternative, we have developed an improved LC–MS based approach for monitoring 30S assembly kinetics using electrospray ionization with time-of-flight (ESI-TOF) analysis of isotope-labeled tryptic peptides. Peptides are both identified and quantitated using the primary LC–MS data, avoiding some of the limitations of the accurate mass time tag MS/MS methods. An alternative quantitation method based on Fourier transform convolution is applied, which allows accurate quantitation of the signal contribution from every isotopomer in the isotope distribution.²⁰ The identification strategy takes advantage of the limited set of proteins in the 30S

ribosome samples and the nitrogen content information encoded in the ¹⁵N-labeled proteins. Other advantages of this approach include high mass accuracy and dynamic range,²¹ efficient ionization, and multiple independent measurements from different peptides for every protein, including S2 and S7. Here, the accuracy and precision of this quantitation method is described, and the method is used for improved measurement of the binding kinetics of r-proteins during in vitro assembly.

MATERIALS AND METHODS

Native 30S subunits were prepared from *E. coli* MRE600 as described previously,²² except that a bead beater (BioSpec) was used to lyse the cells. R-proteins and 16S rRNA were prepared from native 30S subunits, and pulse-chase analysis was performed as previously described¹¹ with minor adjustments. Briefly, 16S RNA was prepared by resuspension in TKM buffer (25 mM Tris-Cl pH 7.5, 30 mM KCl, 20 mM MgCl₂) at 3–5 μ M, heating to 42 °C for 5 min, and slow cooling by dialysis against room temperature TKM buffer at 4 °C for at least 2 h. The r-proteins were thawed on ice and diluted into either TKMD buffer (25 mM Tris-Cl pH 7.5, 1 M KCl, 20 mM MgCl₂) for the pulse, or RB buffer (25 mM Tris-HCl pH 7.5, 330 mM KCl, 20 mM MgCl₂) for the chase. Samples were preheated to 40 °C before beginning the pulse. The pulse was performed with 0.3 μ M 16S RNA, 0.45 μ M ¹⁵N r-proteins in a volume of 600 μ L for 0–40 min. A chase of 225 μ L of 6 μ M ¹⁴N r-proteins was added after the pulse, so that the final ratio of ¹⁴N/¹⁵N protein was 5:1. After the chase was added, the reconstitution was incubated at 40 °C for 40 min. The reconstitutions for the standard curve were performed in a similar fashion, except that preformed mixtures of ¹⁴N and ¹⁵N r-proteins were used for the pulse, the scale was 165 pmol RNA, the samples were incubated at 40 °C for 1 h, and the chase step was omitted.

The 30S subunits from reconstitution reactions were cooled to 4 °C and purified on 10–40% sucrose gradients by ultracentrifugation at 26 000 rpm in a Beckman SW-28 rotor in RB buffer, including 0.5 M NH₄Cl to remove nonspecifically bound excess protein. The gradients were fractionated at 1 mL/min and monitored at 254 nm, and the 30S peak was collected. The proteins and RNA were precipitated with 13% TCA on ice overnight and were pelleted at 14 000g for 25 min. The pellets were washed once with cold 10% TCA followed by cold acetone and dried at room temperature. The pellets were partially resuspended with 10 μ L of 100 mM ammonium bicarbonate. To further dissolve the pellets, they were heated to 65 °C for 2 min and sonicated for 10 min in a room temperature water bath. The r-proteins were reduced with DTT and then alkylated with 40 mM iodoacetamide for 60 min at 37 °C. The r-proteins were then digested with 15–30 ng/ μ L of sequencing grade modified porcine trypsin (Promega) at 37 °C overnight, and the digestion reaction was halted with the addition of formic acid to 0.1%. The samples were centrifuged in a microcentrifuge at 16 000g for 10 min to remove the precipitate before LC–MS analysis.

LC–MS analysis was conducted using an Agilent 1100 series HPLC coupled to an Agilent ESI-TOF mass spectrometer with capillary flow electrospray. The digested r-proteins (5–8 μ L) were injected onto an Agilent Zorbax SB C18 150 mm $\times$ 0.5 mm column.

(13) Bantscheff, M.; Schirle, M.; Sweetman, G.; Rick, J.; Kuster, B. *Anal. Bioanal. Chem.* **2007**, *389*, 1017–1031.

(14) Smith, J. C.; Lambert, J. P.; Elisma, F.; Figeys, D. *Anal. Chem.* **2007**, *79*, 4325–4343.

(15) Huttlin, E. L.; Hegeman, A. D.; Harms, A. C.; Sussman, M. R. *Mol. Cell. Proteomics* **2007**, *6*, 860–881.

(16) Snijders, A. P.; de Koning, B.; Wright, P. C. *J. Proteome Res.* **2007**, *6*, 97–104.

(17) Aebersold, R.; Goodlett, D. R. *Chem. Rev.* **2001**, *101*, 269–295.

(18) Smith, R. D.; Anderson, G. A.; Lipton, M. S.; Pasa-Tolic, L.; Shen, Y.; Conrads, T. P.; Veenstra, T. D.; Udseth, H. R. *Proteomics* **2002**, *2*, 513–523.

(19) Zimmer, J. S.; Monroe, M. E.; Qian, W. J.; Smith, R. D. *Mass Spectrom. Rev.* **2006**, *25*, 450–482.

(20) Sperling, E.; Bunner, A. E.; Sykes, M. T.; Williamson, J. R. *Anal. Chem.* **2008**, *80*, 4906–4917.

(21) Domon, B.; Aebersold, R. *Science* **2006**, *312*, 212–217.

(22) Staehelin, T. M. D. R. *Methods Enzymol.* **1971**, *20*, 449–456.

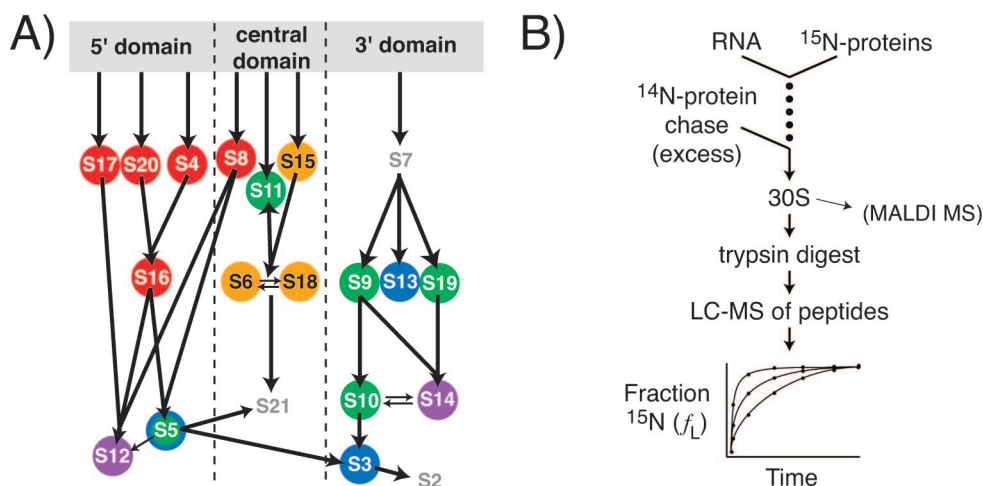


Figure 1. Pulse-chase monitored by quantitative mass spectrometry: (A) equilibrium assembly map of the 30S ribosomal subunit. Arrows represent protein-binding dependencies at equilibrium.^{4,5} Circle colors refer to the apparent protein binding rates previously measured using MALDI-MS under standard reconstitution conditions.¹¹ Red, 20–30 min⁻¹; orange, 8.1–15 min⁻¹; green, 1.2–2.2 min⁻¹; blue, 0.38–0.73 min⁻¹; purple, 0.18–0.26 min⁻¹. S5 is shown in green and blue to represent the binding rates of the unacetylated and acetylated forms, respectively. Positions of the protein symbols along the gray bar correspond to the protein's approximate binding site on the 16S rRNA. Binding rates for S2 and S7 were not determined because those proteins were not visible in the MALDI spectra. (B) Experimental scheme for PC/QMS 30S subunit reconstitution. The previous scheme for PC/QMS involved MALDI-MS of intact ribosomal proteins, whereas the presently described method involves ESI-TOF analysis of tryptic ribosomal peptides.

The mobile phases used were buffer A (H₂O, 0.1% formic acid) and buffer B (acetonitrile, 0.1% formic acid). A gradient was applied to separate the peptides: 5–15% buffer B over 10 min, 15–47% buffer B over 48 min, 47–95% buffer B over 4 min at a flow rate of 7 μ L/min. Data were collected for the m/z range of 100–1300. A separate electrospray probe with a direct infusion of reference ions was used for some samples, which resulted in greater mass accuracy but slightly lower intensity.

RESULTS AND DISCUSSION

The steps for data analysis required for obtaining kinetic r-protein binding curves using PC/QMS with LC-MS of isotopically labeled tryptic peptides are shown in Figure 2. First, a list of experimental peaks is compared to a theoretical digest in order to obtain identities for the ¹⁴N/¹⁵N peptide peak pairs. Then, individual peak pair spectra are fit to a theoretical isotope distribution to obtain the amplitudes of the labeled and unlabeled peaks. Finally, the fraction of labeled intensity as compared to total intensity is plotted against the experimental time points to produce a protein binding progress curve.

A representative LC-MS data set from a sample containing ¹⁴N and ¹⁵N 30S ribosomal peptides is shown in Figure 3A as a grayscale plot. The retention times for the ¹⁴N and ¹⁵N versions of the peptides are essentially identical, with differences typically less than 2 s^{23,24} and many pairs of ¹⁴N/¹⁵N isotopomers are visible as pairs of dark spots. A typical trypsin digest of 30S r-proteins gives rise to several thousand mass peaks, from which ~100–300 ¹⁴N–¹⁵N peak pairs can be confidently identified and quantitated. On the basis of nearly a hundred LC-MS data sets, different subsets of peptides are observed and identified in each ribosomal peptide sample (Figure S1 in the Supporting Information).

Peak Identification. The first step in identifying the peak pairs is to generate a list of experimental peaks. This list is generated from

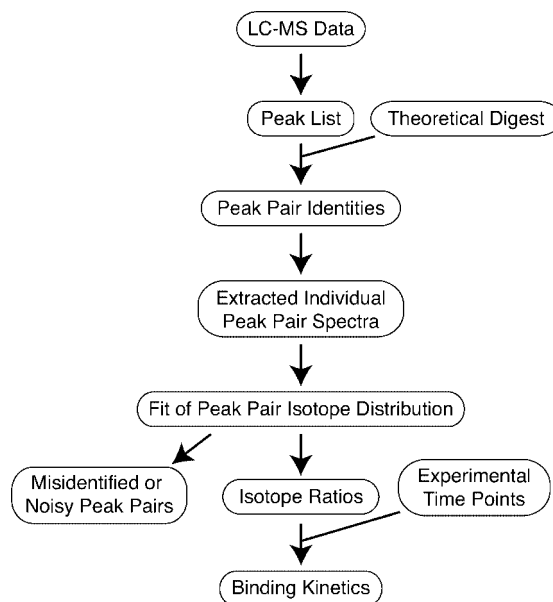


Figure 2. Data analysis flowchart. Peak pairs are identified by comparison with a theoretical digest. Individual peak pair spectra are extracted from the data set and fit using LS-FTC²⁰ to obtain isotope ratios.

the LC-MS data using the Agilent programs MassHunter and MassProfiler, which report the retention time, the m/z of the monoisotopic peak, and the charge state. A theoretical trypsin digest of the set of 20 30S r-proteins was performed using the UCSF MS-Digest program (<http://prospector.ucsf.edu>) to generate possible ¹⁴N m/z values for the 30S ribosomal subproteome, and the m/z values were calculated for the corresponding ¹⁵N peptides based on their molecular formula. The experimental peak list is then compared to the peak list from the theoretical digest of 30S r-proteins to identify the peaks.

The limited set of proteins in the purified 30S subunit samples simplifies the identification process. However, even for a mixture

(23) Snijders, A. P.; de Vos, M. G.; Wright, P. C. *J. Proteome Res.* **2005**, *4*, 578–585.

(24) Zhang, R.; Regnier, F. E. *J. Proteome Res.* **2002**, *1*, 139–147.

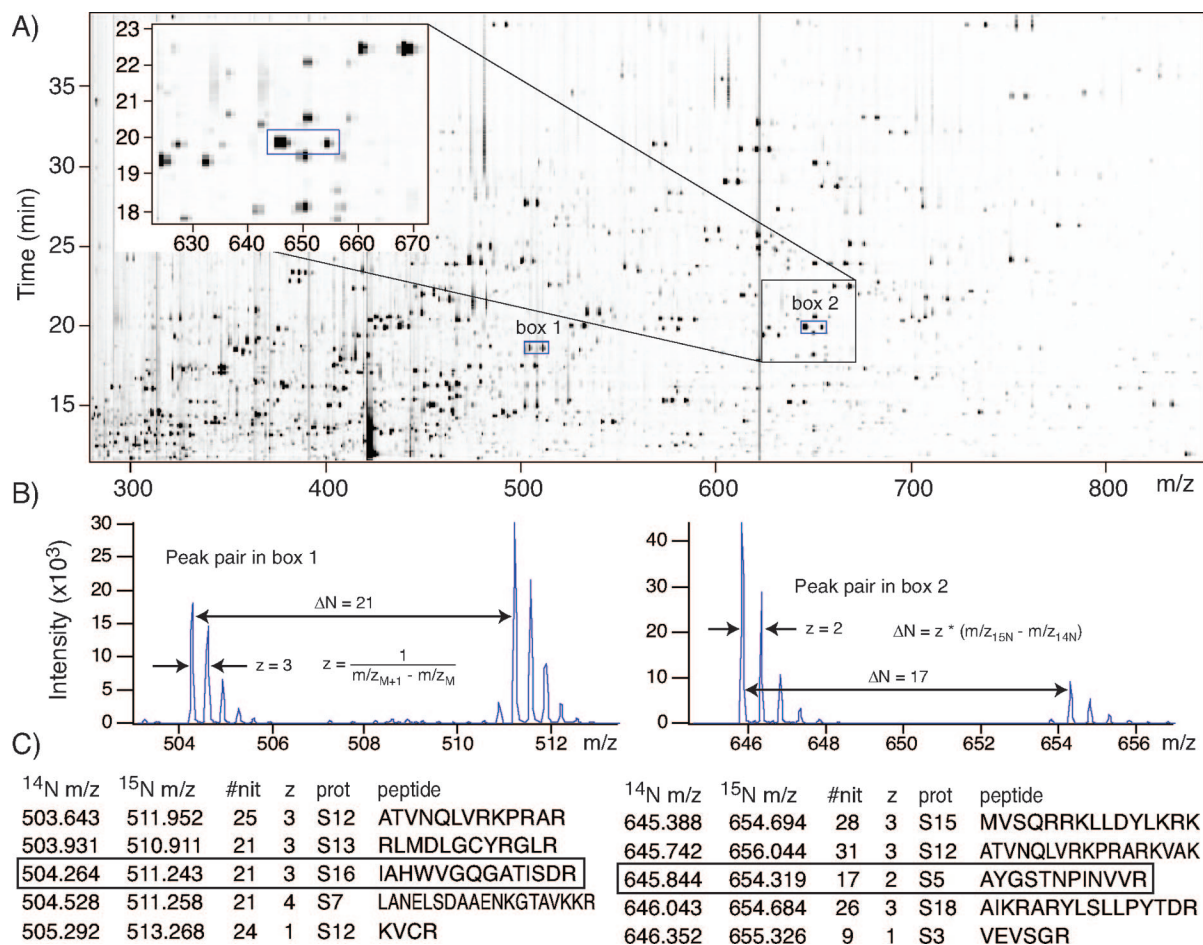


Figure 3. LC-MS data and peptide identification. (A) Image plot showing a typical LC-MS run in low resolution. Intensity is shown in grayscale. Vertical lines at 322, 622 m/z are reference ions. (B) Representative peaks showing isotope distribution. The precursor mass is obtained from the charge state and the m/z value of the ion. (C) Identification of peaks in part B. Several possible identities are listed, with ¹⁵N labeled and unlabeled m/z , number of nitrogens, charge state (z), protein, and peptide sequence. The correct identity is boxed. The ¹⁵N peaks exhibit a small $M - 1$ peak due to the small amount of residual ¹⁴N in the commercially available ¹⁵N ammonium salt.

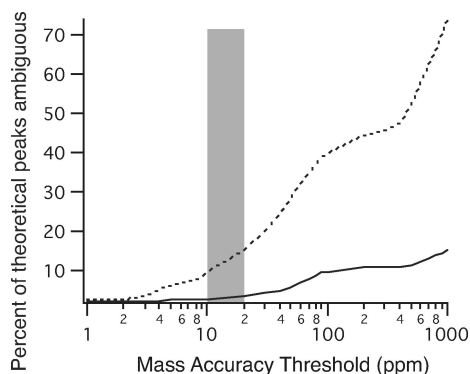


Figure 4. Ambiguity in mass-based peak identification. A theoretical digest ($n \sim 10\,000$) was searched for peptides with unique identities using two criteria based on varied expectations of mass accuracy. The dashed curve shows the level of ambiguity using only the monoisotopic mass for identification, the solid curve using the monoisotopic mass, and the number of nitrogens deduced from a ¹⁴N-¹⁵N peak pair. For the theoretical digest, peptides with up to three missed cleavages were considered, with monoisotopic masses of 200–6000 Da, and charge states $z = 1-4$. Both unmodified cysteines and carbamidomethylated cysteines were included in the theoretical digest, as well as known modifications.³² The gray area represents the maximum errors (10–20 ppm) obtained in a typical ESI-TOF data set.

of peptides from 20 small proteins, the high-resolution precursor mass alone is insufficient to unambiguously identify many peaks due to the complexity of the theoretical digest. Ambiguous identities are greatly reduced by including the nitrogen content as an additional criterion for identification.^{23,25} Features from the experimental peak list are matched as candidate ¹⁴N-¹⁵N peaks pairs, which have very similar retention times, identical charge states, and m/z values that are similar within a few percent. The number of nitrogens in the peptide can be deduced from the experimental difference in m/z between the ¹⁴N peak and the ¹⁵N peak, as shown in Figure 3B. An example comparison between theoretical and experimental peaks is shown in Figure 3C. There is a single possible peptide identity that has the correct mass, charge state, and nitrogen content, indicated by a box. With the use of this strategy, confident identifications can be made for the vast majority of the observed peak pairs. Any peak pairs with ambiguous identities are eliminated from consideration for the quantitative analysis.

A plot of the frequency of ambiguous peak identifications in the 30S protein theoretical digest is shown in Figure 4 as a function of the experimental mass accuracy. Many of the theoretical peaks have similar m/z values, but most of those entries arise from peptides with different numbers of nitrogens.

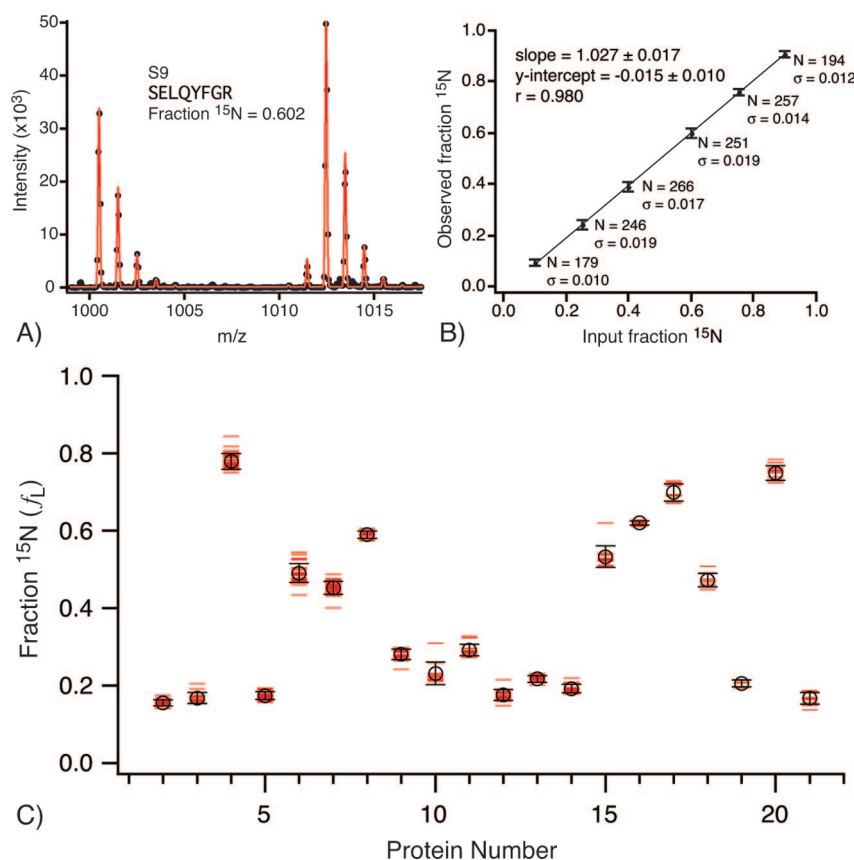


Figure 5. Standard curve demonstrating quantitation. (A) Example LC–MS peak pair showing data (●) and least-squares fit (line). (B) Known mixtures of ¹⁴N and ¹⁵N r-proteins were used in a 30S reconstitution. Peptides from all proteins were pooled to calculate standard deviations and linear fit. (C) Fraction ¹⁵N, $f_L = A_L/(A_U + A_L)$, of r-proteins from 10 s time point from a standard condition reconstitution experiment. Individual peptide measurements are shown as red dashes, and the resulting average f_L values are shown with ○. Error bars are ±1 standard deviation. The average number of observed peaks is 16 ± 7 .

The Agilent ESI-TOF instrument routinely gives a mass accuracy of <20 ppm, which permits unambiguous identification of 96.4% of theoretical peak pairs in the theoretical digest of the 30S ribosomal subunit. For relatively simple peptide mixtures such as the set of 30S ribosomal tryptic peptides, further analysis such as MS/MS is unnecessary for confident identification. Nevertheless, to verify the identities of the peptides observed using ESI-TOF, an unlabeled sample of ribosomal peptides was analyzed both on a Finnigan LTQ ion trap by data-dependent LC–MS/MS and on the Agilent ESI-TOF. Fewer peptides were identified using LC–MS/MS analysis compared to ESI-TOF. Several nonribosomal proteins were also observed in the MS/MS run. Of the 129 peptides observed in both LC–MS/MS and ESI-TOF analysis, 123 of the identities from the ESI-TOF data were confirmed with the MS/MS data. In typical experimental samples, ~98.5% of peaks can be confidently identified. This limited quantity of misidentified peaks does not have a significant impact on the data because the quantitated values from several independent measurements from different peptides are averaged for every protein. In a typical data set from a single time point, between 5 and 19 individual peaks derived from 3 to 11 unique peptides are observed and quantitated per protein.

Isotope Ratio Quantitation. After isotope peak pairs have been assigned an identity, they must be quantitated to obtain isotope ratios and protein binding progress curves for each r-protein. The subspectra of individual peak pairs are extracted

from the larger data set prior to quantitation by summing consecutive scans over ~0.2 min in the region of interest. The amplitudes of the peak pairs for the ¹⁴N and ¹⁵N isotopomer of each peptide are obtained using least-squares Fourier transform convolution (LS-FTC), as described in detail elsewhere.²⁰ Quantitation using LS-FTC to integrate the peak areas is performed on the entire isotope distribution. All of the data points in the peak profile contribute to the integrated intensity, which improves the accuracy of the quantitation. Heavier isotopomers can represent a significant fraction of the total intensity of the peak for larger peptides. The peak fitting procedure provides A_U , the amplitude of the unlabeled peak, and A_L , the amplitude of the ¹⁵N-labeled peak. The fraction of ¹⁵N protein is given by $f_L = A_L/(A_L + A_U)$.

Briefly, the FTC method of generating theoretical isotope distributions was developed by Rockwood²⁶ and was recently adapted for quantitation using least-squares fitting in our laboratory.²⁰ On the basis of the molecular formula of the peptide and the isotope abundances, the theoretical spectrum is first calculated as a conjugate Fourier representation in the frequency domain, then inverse Fourier transformed for comparison to the experimental data. The formulas for calculating these spectra allow for variations in the baseline, peak width, mass accuracy, and amplitude of experimental peaks. The enrichment of ¹⁵N in the

(25) Nelson, C. J.; Huttlin, E. L.; Hegeman, A. D.; Harms, A. C.; Sussman, M. R. *Proteomics* **2007**, *7*, 1279–1292.

(26) Rockwood, A. L.; Van Orden, S. L.; Smith, R. D. *Anal. Chem.* **1995**, *67*, 2699–2704.

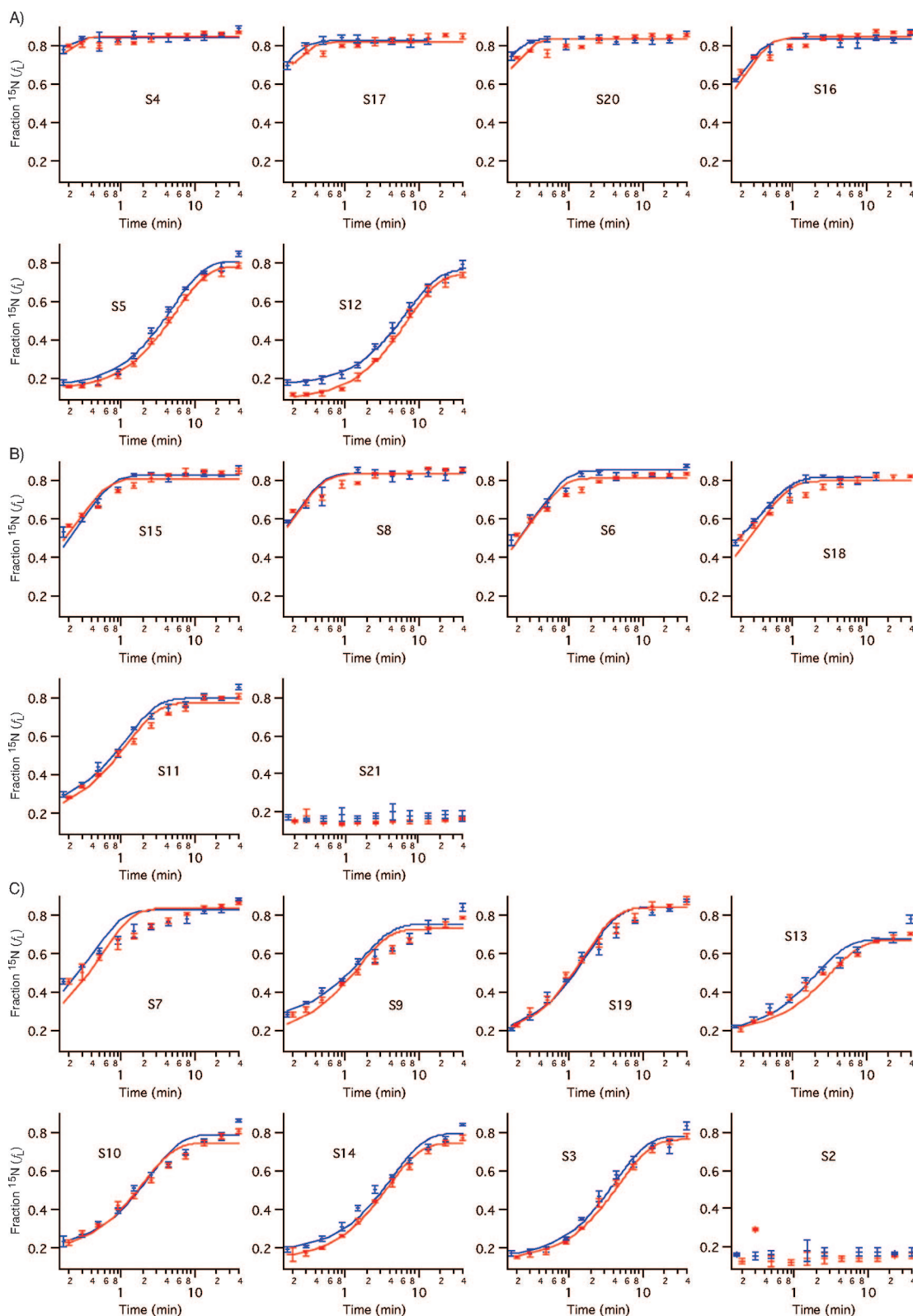


Figure 6. Protein binding progress curves measured by PC/QMS with LC-ESI-TOF. Curves from two experiments are shown, with curves fit to single exponentials (red and blue). Observed binding rates are reported in Table S1 in the Supporting Information. (A) 5' domain proteins, (B) central domain proteins, and (C) 3' domain proteins. S7 and S9 show multiphasic binding kinetics.

labeled peptides is held fixed during the fitting at 99.3%, which was the empirically observed ^{15}N enrichment in the samples, although this quantitation approach can be applied to peptides labeled in any proportion. An example extracted individual peak-pair spectrum with the least-squares fit is shown in Figure 5A.

For large data sets of several hundred peaks, an automated fitting procedure is applied to the set of extracted individual peak pair spectra. Although the fitting procedure is robust, the fits for ~15–25% of extracted spectra fail to converge due to noisy data or overlapping peaks interfering with the fit. These unconverged

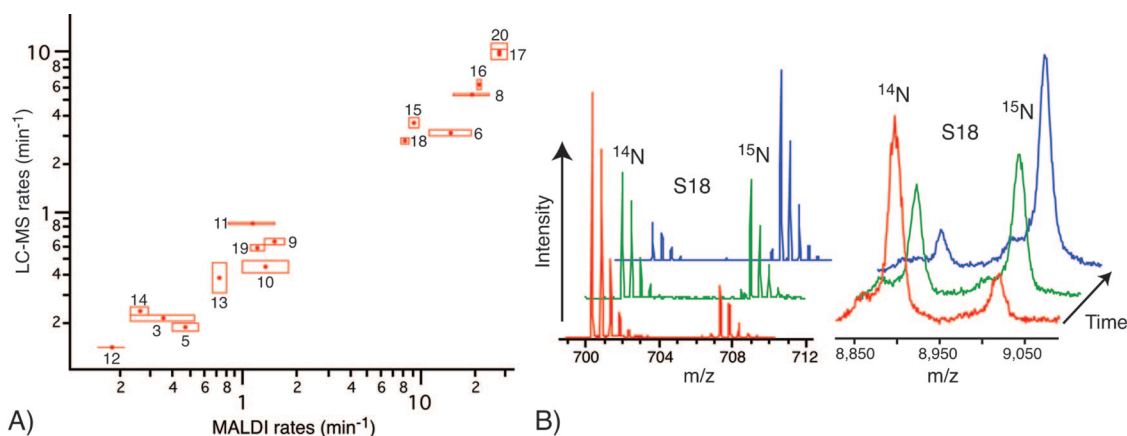


Figure 7. Progress curves and comparison with MALDI method. (A) Comparison of rates obtained using MALDI and LC–MS. Boxes indicate the range of values in two experiments. S2 and S7 were not observed in the MALDI spectra. S4 is not shown because the binding transition was too fast to accurately observe. (B) S18 peaks from MALDI and LC–MS. The left panel shows an S18 peptide observed with LC–MS. The right panel shows protein S18 observed with MALDI.

fits are recognized by their nonsensical parameters, and the corresponding extracted spectra are automatically eliminated from further consideration. The fitted spectra are visually inspected, and peak pairs that are overlapped or incorrectly paired are also eliminated from the analysis. The fraction labeled (f_i) values for each protein are obtained by averaging the values from the set of peptides observed, and the standard deviation of each set is used as an estimate of the quantitation error for each protein.

In order to demonstrate the accuracy of the quantitation procedure, 30S reconstitutions were performed with mixtures of ^{14}N and ^{15}N r-proteins in five known proportions. The fraction ^{15}N data are shown in Figure 5B for each data set, plotted against the input ^{15}N fraction. The correlation between the observed fraction ^{15}N signal intensity and the input fraction ^{15}N protein is excellent. The standard deviations for the measured fraction labeled values are all less than 0.02, which is an improvement over errors of 0.03–0.04, which were seen with similar experiments with MALDI of whole proteins.¹¹

A typical time point from a pulse-chase experiment with proteins bound to varying extents is shown in Figure 5C. The small errors in this experimental data illustrate both the precision of the quantitation and the accuracy of the identification strategy. Since peptides from different proteins have different isotope ratios due to their different binding kinetics, inaccurately identified peptides would increase the errors in fraction labeled values. Occasional outliers are usually due to misidentified peaks and are eliminated from further analysis. This excellent quantitation strategy enables measurement of more accurate kinetic curves that better define multiphasic kinetics.

Assembly Kinetics. The 30S in vitro assembly kinetics were examined under standard conditions (40 °C) using ESI-TOF-PC/QMS of tryptic peptides. Stable-isotope labeled proteins were incubated with 16S rRNA for varied amounts of time before the assembly reaction was chased with an excess of unlabeled protein. Peak pairs from 12 time points between 0 and 40 min were identified and their isotope ratios quantitated using LS-FTC. Individual peptides from the same protein have very similar binding kinetics, as shown in Figure S2 in the Supporting Information. The average fraction labeled (f_i) values for all

peptides from each protein are plotted against experimental time points for two independent experiments to show r-protein binding progress curves in Figure 6. These progress curves were fit to a single exponential using the standard deviation of the fraction labeled for the set of peptides observed at each time point to weight the fit. The data for S7 clearly do not fit the single exponential curve, indicating multiphasic kinetics. This is the first time that data from all 20 30S r-proteins have been recorded in a single experiment, as S7 and S2 were not observable in previous kinetic experiments using MALDI.

The protein-binding rates observed were generally similar to those previously reported.¹¹ Figure 7A shows a log–log plot of the rates observed using MALDI and the rates reported here, with LC–MS. The pattern of the rates is the same between the two methods, and most LC–MS k_{obs} values (Table S1 in the Supporting Information) are within a factor of 2 of their counterparts in the MALDI data set. Figure 7B illustrates the difference in the primary data acquired by these two methods, where the S18 peak for several time points as observed in MALDI and an S18 peptide from similar time points observed with LC–MS are shown. There is a dramatic improvement in signal-to-noise ratio for the LC–MS method, which contributes to the superior quantitation.

S7 is the sole primary binding protein in the 3' domain and nucleates the 3' domain assembly,²⁷ which is the slowest domain to assemble in vitro at all temperatures. The overall rate is reported here for the first time and is consistent with previous qualitative findings.⁶ The multiphasic kinetics of S7 are likely due to the presence of several alternate pre-30S particles in different regions of the assembly landscape binding S7 at different rates. This observation is consistent with in vitro assembly proceeding through multiple parallel pathways.^{11,12} S9, a 3' domain protein dependent on S7 at equilibrium, also shows multiphasic kinetics (Figure S3 in the Supporting Information). Multiphasic kinetics have previously been observed for S8 and S15 at 15 °C conditions using MALDI-PC/QMS.¹¹

Binding rates for proteins S2 and S21 could not be determined because those proteins exhibited low values of f_i for all time points.¹¹ The observed value of f_i for S2 and S21 was ap-

(27) Nowotny, V.; Nierhaus, K. H. *Biochemistry* **1988**, *27*, 7051–7055.

proximately equal to the fraction of labeled protein in the reconstitution after the chase was added (1:5), suggesting that the proteins present in the final 30S subunits bound during the chase. However, it is unlikely that these proteins were entirely unbound during the pulse, since S2 and most other 30S r-proteins loosely associate with their binding sites very quickly, as shown by recent time-resolved hydroxyl radical footprinting experiments.²⁸ For technical reasons, these experiments provided only limited information about the binding kinetics of S2 and S21. In our PC/QMS experiments, it is likely that S2 and S21 have a fast off-rate in reconstitution buffer, and bound protein exchanges with excess protein in solution during the chase. If these proteins did bind transiently, it might be expected that some S2 and S21 would also be lost during sucrose gradient purification under nonequilibrium conditions. The stoichiometries of S2 and S21 are somewhat less than that of other ribosomal proteins (Figure S4 in the Supporting Information). We also observed S2 and S21 exchange from native 30S subunits under reconstitution conditions (Figure S5 in the Supporting Information), which is consistent with previous reports suggesting exchange during or after 30S subunit purification.^{29–31} Exchange is a fundamental limitation to any pulse-chase strategy, but fortunately, the exchange rates

of 18 of 20 proteins are sufficiently slow to permit measurement of their binding kinetics using PC/QMS.

CONCLUSIONS

We have developed an improved general LC–MS-based method for monitoring self-assembly in vitro. A high degree of accuracy and precision was achieved in quantitation of relative amounts of ¹⁴N and ¹⁵N peptides. This PC/QMS method could be applied to any self-assembling particle that can be reconstituted and separated from individual components. The binding rates of 18 of the 20 30S r-proteins have been measured for the first time in a single experiment using PC/QMS. A complete progress curve for S7 shows multiphasic kinetics, providing evidence for multiple pre-30S intermediate species assembling through different pathways.

ACKNOWLEDGMENT

The authors thank Dr. Michael T. Sykes and Dr. Zahra Shajani for critical comments on the manuscript. The authors also thank Dr. Michael T. Sykes for programs that assist with data analysis. This work was supported by NIH Grant R37-GM53757 and the Skaggs Institute for Chemical Biology.

SUPPORTING INFORMATION AVAILABLE

Additional information as noted in text. This material is available free of charge via the Internet at <http://pubs.acs.org>.

Received for review September 28, 2008. Accepted October 28, 2008.

AC8020505

(28) Adilakshmi, T.; Bellur, D. L.; Woodson, S. A. *Nature* **2008**, *455*, 1268–1272.

(29) Ulbrich, B.; Nierhaus, K. H. *Eur. J. Biochem.* **1975**, *57*, 49–54.

(30) Subramanian, A. R.; van Duin, J. *Mol. Gen. Genet.* **1977**, *158*, 1–9.

(31) Robertson, W. R.; Dowsett, S. J.; Hardy, S. J. *Mol. Gen. Genet.* **1977**, *157*, 205–214.

(32) Arnold, R. J.; Reilly, J. P. *Anal. Biochem.* **1999**, *269*, 105–112.