Analytical Biochemistry 560 (2018) 39–49

Contents lists available at ScienceDirect

Analytical Biochemistry

journal homepage: www.elsevier.com/locate/yabio

Re-evaluation of the rabbit myosin protein standard used to create the empirical statistical model for decoy library searching

Jaimie Dufresne, Angelique Florentinus-Mefailoski, Pei-Hong Zhu, Peter Bowden, John G. Marshall*

Ryerson Analytical Biochemistry Laboratory (RABL), Kerr Hall East Rm 332b, Department of Chemistry and Biology, Faculty of Science, Ryerson University, Canada

ARTICLE INFO

Keywords:
Tryptic peptides
Random and independent sampling
LC-ESI-MS/MS
Chi square
ANOVA
Type I error rate

ABSTRACT

A Rabbit myosin standard, like that used to create the empirical statistical model, was randomly and independently sampled by liquid chromatography micro electrospray ionization and tandem mass spectrometry (LC-ESI-MS/MS) with a linear quadrupole ion trap. The rabbit myosin protein standard appeared pure by SDS-PAGE and CBBR staining but showed many other proteins by silver staining. The LC-MS intensity from myosin and IgG samples were above the 99% safe limit of detection and quantification computed from 36 blank LC-ESI-MS/MS runs. The myosin contained ≤406 Gene Symbols, open reading frames or loci where 79 protein types showed ≥3 peptides from X!TANDEM. Myosins, actin, troponin, other proteins showed 95%–100% homology between the rabbit versus the human decoy library. The myosin protein complex from STRING was true positive compared to random or noise spectra MS/MS with a low type I error (p-value) and low FDR (q-value) computed in R. SDS-PAGE, Western blot, comparison to random and noise MS/MS spectra, X!TANDEM p-values, FDR corrected q-values, and STRING all agreed that the error rate of LC-ESI-MS/MS with a quadrupole ion trap is far below that assumed *a priori* by the design of the empirical statistical model for decoy library searching.

Introduction

The proteins of a complex Eukarytic sample may by identified and quantified by liquid chromatography, micro electrospray ionization, and tandem mass spectrometry (LC-ESI-MS/MS) with a simple ion trap mass spectrometer [1–3]. However, the Empirical Statistical Model that is the basis for Decoy library Searching indicated that LC-ESI-MS/MS may suffer from a high False Discovery Rate (FDR) [4,5]. By fusing the sequence for rabbit myosin to a human protein library, the Empirical Statistical Model for decoy library searching assumed that the correlations to rabbit Myosin were true positive, that all other correlations were *necessarily* false positive, that a score distribution may resolve true positive and false positive peptides, and that the proteins in the Myosin standard are not shared by the decoy human library [4,5]. Here the central assumptions of the empirical statistical model; that the myosin standard contains only one true positive protein; that there are no homologous proteins shared between rabbits and humans; and that individual peptide p-value distributions resolve true positive results from false positive, were all shown to be incorrect. Myosin may be isolated by selective extractions and precipitations to a preparation that is apparently homogenous by SDS-PAGE Coomasie brilliant blue R (CBBR) staining [6,7]. Similarly, IgG may be purified by affinity chromatography [8]. Affinity purified human IgG contains constant chains that are well characterized and show low sequence variation, but also contains variable domains with some regions of near random polypeptides [9,10]. Myosin can be detected by Coomasie blue staining that is sensitive to about ~1 μg [11]. In contrast, liquid chromatography electrospray ionization and tandem mass spectrometry (LC-ESI-MS/MS) is a sensitive to ≤ 1 ng on column [12–14]. The error rate relative to random MS/MS spectra has previously been estimated in a 3D Paul ion trap that showed that authentic MS/MS spectra, were concentrated on a small set of true positive proteins with a high frequency, but in contrast, false positive random spectra or electrospray source noise (chemical and physical noise) were randomly spread across the peptide database [15–17].

The computation of p-values from the fit of the MS/MS spectra [18], comparison of the frequency distribution of the results to that of a reference library [19–21], the agreement between independent instruments and groups [22], excellent agreement between fully tryptic versus no enzyme searches [19,22,23], and computing the

Abbreviation: aa, amino acids; RSG, random MS/MS spectra generator

* Corresponding author. Department of Chemistry and Biology, Ryerson University, 350 Victoria Street, M5B 2K3, Toronto, Canada.
E-mail address: 4marshal@ryerson.ca (J.G. Marshall).

https://doi.org/10.1016/j.ab.2018.08.025
Received 12 December 2017; Received in revised form 22 August 2018; Accepted 28 August 2018
Available online 29 August 2018
0003-2697/ © 2018 Published by Elsevier Inc.

permutations [24] all indicate that LC-ESI-MS/MS shows a low error rate as confirmed by biochemical and immunological methods as well as specific drugs or siRNA [25–28]. A study with 300,000 synthetic peptides, or pure viral cultures [29], demonstrated that the observed MS/MS spectra closely match the predicted [30] and so correlation of MS/MS spectra should show a low error rate as observed herein. The False Discovery Rate of protein p-values estimated by a correlation algorithm such as X!TANDEM can be corrected by the method of Benjamini and Hochberg to generate the corresponding q-value (FDR) [23,31]. Classical frequency based statistical methods including random and independent sampling, comparison to random expectation, Chi square test of peptide counts, transformation of ion intensity to a Gaussian distribution and normality testing [32] for ANOVA [25,33], may be applied prior to correlation of MS/MS spectra, after making labeled derivatives, or after peptide identification by LC-ESI-MS/MS [25,33–37].

Here the same type of rabbit Myosin standard used to create the Empirical Statistical Model [4] was randomly and independently sampled by LC-ESI-MS/MS alongside the controls of IgG, laboratory dust, source noise MS/MS spectra from blank solvent, or computer random MS/MS spectra. As a traditional statistical control, the well-established algorithm for generating random numbers from Park and Miller [38] was harnessed to create a random MS/MS spectra generator (RSG) [16] that serves, along with noise from blank runs on naïve columns, as a null random model of false positive results. The peptide p-values provided by X!TANDEM and stored in SQL Server were computed in the R statistical system to provide the cumulative protein p-value and corresponding FDR corrected q-values [39] for each protein or Gene Symbols. It was clearly demonstrated that while Myosin heavy chain binds CBBR efficiently and appears pure at low sample loadings, the myosin standard actually contains hundreds of true positive proteins in marked contrast to the central assumption of the Empirical Statistical Model that is the basis of computing FDR by decoy library searching [4]. Comparison to the frequency distribution of random MS/MS correlation is the gold standard and the best statistical practice [15–17] where no p-value is available.

Here the for the first time results of the myosin standard from the linear quadrupole ion trap are compared to that of random expectation by two methods that clearly demonstrated the error rate of LC-ESI-MS/MS with a linear quadrupole ion trap is much lower than previously assumed [4]. Also for the first time, the tryptic peptides from the Eukaryotic myosin protein standard were randomly and independently sampled on a linear quadrupole ion trap above the 99% safe limit of detection and quantification compared to 36 blank source noise LC-ESI-MS/MS recordings (blank mean intensity + 3 standard deviations) with quantification of Gaussian $\log_{10}$ intensity values by ANOVA [40]. In agreement with results from the 3D Paul trap [15,16,34], SDS-PAGE, Western blot, comparison to random MS/MS peptide observation frequency, MS/MS spectra p-values with FDR corrected q-values, and STRING all confirm the myosin standard is a protein complex, and not a single pure protein, with demonstrated low levels of type I error in primary structural identity. The peptide frequency and $\log_{10}$ peptide intensity values from micro electrospray LC-ESI-MS/MS with a sensitive and robust linear quadrupole ion trap can be identified and safely quantified with high confidence using the classical statistical techniques such as random sampling, Chi Square analysis and ANOVA.

Material and methods

The 1100 HPLC system was from Agilent (Santa Clara, CA, USA). The LTQ XL linear ion traps were obtained from Thermo Electron Corporation (Waltham, MA,USA). The random spectra generator was slightly modified from that provided by Zhu et al. [15]. The C18 with 5 micron particle size and 300 Angstrom pore size was from Agilent (Zorbax 300 SB-C18 5-micron). The HPLC grade water and acetonitrile was from Caledon laboratories and the formic acid from Fluka (Georgetown, ON, Canada). Sequencing grade trypsin was from Roche (Basel, Switzerland). The C18 Zip Tips were from Millipore (Billerica, MA, USA). The beta Actin Monoclonal Antibody (AC-15) was obtained from Thermo Fisher Scientific (Waltham, MA USA) and the Donkey anti mouse secondary antibody was obtained from Jackson Immuno Research (West Grove, PA, USA). The bulk human IgG and rabbit myosin standards were obtained from Sigma Aldrich and except where indicated other reagents, buffers and salts were obtained from Sigma Aldrich (St Louis, MO USA). Dust samples were collected from Kerr Hall East room 215 of Ryerson University.

SDS-PAGE and western blot

The rabbit myosin standard (50 μg) was separated over 1.5 mm 9% polyacrylamide tris gels with tricine electrophoresis buffer [41] and either stained in 1% Coomassie Brilliant Blue R (CBBR) in 40% methanol and 5% acetic acid, or reduced in glutaraldehyde and stained by $AgNO_3$ [42] or transferred to PVDF and probed with anti actin antibody (1/500) and washed before detection with anti-mouse secondary antibody (1/2000) with detection by ECL [43].

Calibration, tuning, conditioning and sensitivity

The linear ion trap instrument [44] was cleaned and calibrated with the manufacturers' standards. The instrument was tuned and mass accuracy confirmed with the direct infusion of angiotensin and glu-fibrinogen immediately prior to the experiments. The columns were preconditioned and quality control tested with a digested mixture of alcohol dehydrogenase, cytochrome *c* and glycogen phosphorylase to confirm the system was working normally versus historical benchmarks [45]. The micro electrospray voltage was typically ~4.5 kV with ~10 L N_2 per minute and the transfer capillary was heated to 200 °C. The instrument was set at 20 milli seconds for the wide scan (350–2000 *m/z*) to randomly and independently sample the 5 most intense peptides eluting without replacement every few moments for up to 200 milli seconds to fill the trap with up to 250,000 ions, and to continue collecting MS/MS spectra for up to 4 independent MS/MS fragment scans that were averaged. The sensitivity of the system was tested with a BSA digest on the conditioned columns that showed a sensitivity for automatic identification by SEQUEST to ~1 fmol on column.

Liquid chromatography electrospray ionization & tandem mass spectrometry

Three LC-ESI-MS/MS systems were fitted with naive 300-micron ID silica C18 columns Agilent (Zorbax 300 SB-C18 5-micron) with 20 μL sample loops and blank samples of 5% formic acid and 5% acetonitrile were injected into a 2 μL flow of 5% acetonitrile and 0.1% acetic acid from the micro electrospray source. The gradient of acetonitrile was commenced after 12 min from 15% to 45% ACN over the course of 60 min and then to 65% ACN over 30 min, cleaning at 65% for 5 min before returning to 5% ACN. A total of 36 replicate blank recordings with some 75,269 source noise MS/MS spectra of any intensity were recorded and analyzed without excluding precursor ions of less than 10,000 counts. For the comparison of no enzyme to trypsic peptide fitting, additional myosin samples were analyzed with 150-micron ID silica by nano spray.

IgG, myosin and dust tryptic protein digestion and analysis

Precursor values of greater than ~10,000 (E4) arbitrary counts were accepted for the IgG, myosin and dust samples in order to avoid the collection of chemical or physical source noise spectra in these treatments. Dust immediately proximal to the LC-ESI-MS/MS instruments were sampled at 4 locations in the Ryerson Analytical Biochemistry Laboratory (RABL). The IgG, Myosin and dust were digested with a 1/100 ratio (g/g) of trypsin for 8 h in 200 mM urea and

Download English Version:

https://daneshyari.com/en/article/10147256

Download Persian Version:

https://daneshyari.com/article/10147256

Daneshyari.com