

Molecular Detection of Targeted Major Histocompatibility Complex I-Bound Peptides Using a Probabilistic Measure and Nanospray MS³ on a Hybrid Quadrupole-Linear Ion Trap

Bruce Reinhold,^{*,†,‡} Derin B. Keskin,[‡] and Ellis L. Reinherz^{†,‡}

Cancer Vaccine Center and Laboratory of Immunobiology, Department of Medical Oncology, Dana-Farber Cancer Institute, Harvard Medical School, Boston, Massachusetts, United States

A nanospray MS³ method deployed on a quadrupole linear ion trap hybrid can detect targeted peptides with high dynamic range and high sensitivity from complex mixtures without separations. The method uses a recognition algorithm that is a modification of the relative (Kullback–Leibler, KL) entropy characterization of probabilistic distance to detect if reference MS³ fragmentation patterns are components of acquired MS³ spectra. The recognition reflects the probabilistic structure of physical MS measurements unlike the Euclidean or inner product metrics widely used for comparing spectra. It capably handles spectra with a significant chemical ion background in contrast to the Euclidean metric or the direct relative entropy. The full nanospray MS³ method allows both the detection and quantitation of targets without the need to obtain isotopically labeled standards. By avoiding chromatographic separations and its associated surface losses, the detection can be applied to complex samples on a very limited material scale. The methodology is illustrated by applications to the medically important problem of detecting targeted major histocompatibility complex (MHC) I associated peptides extracted from limited cell numbers.

Mass spectrometry (MS) can operate in a discovery or a detection mode. Acquiring MS and “data-dependent” MS/MS spectra in a linear ion trap (LIT) to identify peptide components of a complex sample exemplifies discovery mode whereas acquiring parent/fragment multiple reaction monitoring (MRM) transitions in a tandem quadrupole is operating in a detection format. Although the distinction between discovery and detection is not rigorous (discovery within the confines of a database merges into detection), the practical sense is that discovery analysis returns a list of the components in a sample whereas detection analysis targets specific molecules and is indifferent to the overall composition. Recent developments in MS for “-omics” scale

discovery analyses have been impressive. However, important problems in analytical biology are not effectively addressed by discovery technologies and methods. These limitations relate primarily to the ability to detect and quantitate a few targeted components that are a small fraction of a sample that is itself small. The challenge is to address both the high dynamic range required of these analyses and to maintain high absolute sensitivity. Of course, dynamic range and sensitivity are also desired in discovery MS. It is targeting that provides the most dramatic opportunities to enhance detection. Analytical problems characterized by the triad of small sample scale, high degree of complexity, and a restricted set of targets are identified here as molecular detection problems. A targeted detection methodology, denoted as MS³ Poisson detection, acquires and analyzes MS³ spectra using a hybrid quadrupole-LIT instrument and a probabilistic measure. This study develops a formal framework for probabilistic MS detection and illustrates the method by application to major histocompatibility complex (MHC) I peptides.

MHC I molecules (termed human leukocyte antigens, HLA, in humans) bind normal self-peptides as well as pathologic peptides derived from infectious organisms, tumors, or stress-related proteins.^{1–5} Pathologic peptides, bound to MHC I, are arrayed on cell surfaces for immune recognition. They mark the cell as infected or transformed and target cell destruction by cytolytic T lymphocytes (CTLs).^{6,7} Consequently, detection and quantitation of immunologically relevant peptides associated with MHC I molecules on defined cell populations are enormously significant challenges, both in basic immunology research and for medical applications. The vast majority of peptides displayed on presenting cells are not pathologic and reflect normal cellular processes. Generally, some effort is invested in identifying targets of interest from this sea of uninteresting, nonpathologic peptides. For example, to target peptides derived from intracellular pathogens, one combines knowledge of the pathogen’s proteome with

* Corresponding author. Address: Harvard Medical School, Cancer Vaccine Center, Dana Farber Cancer Institute, HIM 422, 77 Ave Louis Pasteur, Boston, Massachusetts 02115, U.S.A. Phone: 617-632-3121. Fax: 617-632-3351. E-mail: bruce_reinhold@dfci.harvard.edu.

[†] Cancer Vaccine Center.

[‡] Laboratory of Immunobiology, Department of Medical Oncology.

- (1) Groothuis, T.; Neeffjes, J. *Curr. Top. Microbiol. Immunol.* **2005**, *300*, 127–148.
- (2) Shastri, N.; Cardinaud, S.; Schwab, S. R.; Serwold, T.; Kunisawa, J. *Immunol. Rev.* **2005**, *207*, 31–41.
- (3) Kloetzel, P. M. *Biochim. Biophys. Acta* **2004**, *1695*, 225–233.
- (4) Lehner, P. J.; Cresswell, P. *Curr. Opin. Immunol.* **2004**, *16*, 82–89.
- (5) Yewdell, J. W.; Reits, E.; Neeffjes, J. *Nat. Rev. Immunol.* **2003**, *3*, 952–961.
- (6) Bleackley, R. C. *Biochem. Cell Biol.* **2005**, *83*, 747–751.
- (7) Wang, J. H.; Reinherz, E. L. *Mol. Immunol.* **2002**, *38*, 1039–1049.

well-known constraints on the motif for a peptide to bind to a given MHC I molecule.^{8–10}

Nanospray MS³ Poisson detection focuses on efficient ionization, reduced surface losses, high operational duty cycle, detection specificity through two MS selection stages, and a probabilistic algorithm for pattern recognition. It is targeted, requiring prior knowledge of molecular m/z and MS² and MS³ fragmentation patterns. The detection is limited in the number of analytes monitored as detection of any one analyte is at the expense of any other. However, for a few targets, it provides a uniquely sensitive method for the detection/quantitation of trace components from small, but very complex, samples. A discussion of probabilistic detection will be followed with experimental analyses of targeted MHC I peptides by nanospray MS³ using the QTrap 4000.

THEORETICAL

The following section considers the analytical problem of detecting a target in a matrix by ion fragmentation when the selected m/z window contains the target along with multiple other ion species. In this application, the fragmentation spectrum of the target is known (e.g., previously measured) but chemical background from coselected ions has significant overlap with the target spectrum. MS measurement is formalized as sampling a random process characterized by an underlying distribution (the reference spectrum), and detection is associated with the probability that a measured spectrum could arise from sampling this process. However, calculating this probability is not an effective detection method when chemical noise overlaps with a subset of the reference peaks. To address chemical noise, the probability measure needs to be restricted, and to explain this, some formal structure needs to be developed.

A spectrum is represented as a finite sampling of an independent, identically distributed (in time) random process defined on a finite outcome space D , where the elements of D correspond to the m/z data points established by the acquisition hardware. A reference spectrum is asymptotically characterized by the normalized measure or probability distribution $\vec{P} = (p_1, p_2, \dots, p_{O(D)})$ where $\sum_{j=1}^{O(D)} p_j = 1$. Sampling of this process can be described with a Poisson model where each data channel j in D is associated with an ion arrival rate α_j . The arrival rates α_j can be related to a normalized distribution $\vec{P} = (p_1, p_2, \dots, p_{O(D)})$ on D by introducing a unit arrival period T such that $\sum_{j=1}^{O(D)} T\alpha_j = 1$ or $p_j = T\alpha_j$. To relate the Poisson model to a measured spectrum, suppose the random process is sampled for a period NT and one counts the event distribution $\{n_1, n_2, \dots, n_{O(D)}\}$. For a Poisson process on a single channel characterized by an arrival rate α and a sampling period NT , the probability of measuring n events is $P(n) = (1)/(n!)(\alpha NT)^n e^{-\alpha NT}$. As event arrivals are independent and translating arrival rates into underlying probabilities by $p_j = T\alpha_j$, the probability of the event distribution $\{n_1, n_2, \dots, n_{O(D)}\}$ after an NT sampling period is given by

$$P\{n_1, n_2, \dots, n_{O(D)}\} = \prod_{j=1}^{O(D)} \frac{1}{n_j!} (Np_j)^{n_j} e^{-Np_j} \quad (1)$$

Applying Stirling's approximation for the factorial $n! = (2\pi n)^{1/2} n^n e^{-n}$ and introducing the normalized measure \vec{v} by $N\vec{v} = (n_1, n_2, \dots, n_{O(D)})$, we have

$$\ln [P(\vec{N} \parallel \vec{p})] = -\frac{O(D)}{2} \ln(2\pi N) - \frac{1}{2} \sum_{j=1}^{O(D)} \ln(v_j) - N \sum_{j=1}^{O(D)} v_j \ln\left(\frac{v_j}{p_j}\right) \quad (2)$$

Identifying the relative or Kullback–Leibler (KL) entropy by

$$I_2(\vec{v}, \vec{p}) = \sum_{j=1}^{O(D)} v_j \ln \frac{v_j}{p_j} \quad (3)$$

the result (2) states that, given a reference spectrum and a measurement of that spectrum (without chemical noise), the probabilistic “difference” between the spectra is asymptotically an exponentially decaying function of the number of events times the relative entropy.

Calculating the probabilistic distance between a measured and reference spectrum is not an effective detection algorithm. Chemical noise as ion fragments from coselected molecular components are expected in the MSⁿ spectra of complex mixtures, but any data channel where finite chemical background events are measured and where reference events are not expected (specifically where $v_j \gg p_j$) results in a high entropy cost in (3) and a low probability in (2). This is not wanted; reflecting the expectation of chemical noise, low detection probabilities are to be associated only with $v_j \ll p_j$. Instead of representing the measured events as a finite sampling of the reference distribution, we will consider detection by calculating the probability that M reference events could be contained in the measured events and focus on the decrease in probability as M increases. Restated in the context of MS data, some peaks in the measured spectrum will limit the amount of the target that could be present; with a limiting amount of target established, other peaks in the measured spectrum will have too many events and these additional events will be identified as chemical noise.

Formally, let M be an integer representing a potential number of target events, i.e., events distributed by \vec{p} . For fixed $N\vec{v}$ and \vec{p} , every M splits the outcome space D into D_M (on which $Nv_j < Mp_j$) and its complement ($D - D_M$). Events in the complement correspond to channels with too many events for M total reference events. That is, they are peaks obscured by chemical noise and are ignored. One combines (2) with the M -dependent condition that $Nv_j < Mp_j$ (for each ion peak j) to create the strictly decreasing function $P(M)$.

Random process sampling with chemical noise calculates a probability $P(M)$ that the measured data supports M events of the reference distribution. $P(M)$ is a scoring function whose utility is relative, that is, it is used to compare different patterns. For example, one could maximize $P(M)$ over a set of reference distributions as a method of finding an optimal distribution to represent that data. A measure of confidence or significance in the optimization is the uniqueness of the fit, i.e., comparing the optimal fit with other nearly optimal fits from the reference set.

(8) Madden, D. R. *Annu. Rev. Immunol.* **1995**, *13*, 587.

(9) Stern, L. J.; Wiley, D. C. *Structure* **1994**, *2*, 245.

(10) Wang, J.-H.; Reinherz, E. L. *Curr. Opin. Struct. Biol.* **2000**, *10*, 656.

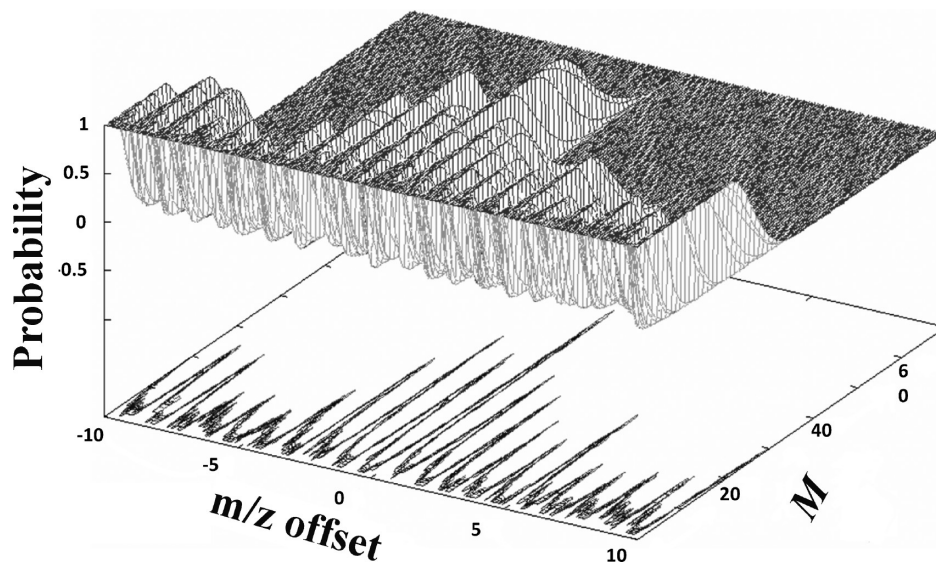


Figure 1. Probabilistic MS detection with chemical noise combines a formal calculation of a probability of target events (the M axis) plotted against an m/z translation to identify uniqueness in the M -dependent probability (m/z offset axis). The resulting probability surface is typically represented by a single contour (fixed probability) with the relative amplitude of the 0 offset peak of the contour plot related to the degree of confidence in the detection.

For detection of a target MS³ reference pattern, there is typically no fixed set over which to optimize but there remains the concept of the significance of $P(M)$ associated with its uniqueness. The implementation we have employed here is to generate other patterns from the reference spectrum by translation in m/z space and calculate and compare the probabilities for the translated spectra with the untranslated one. Translation by m/z is also done in the correlation function method for spectral detection¹¹ and for the same underlying reason, that significance is associated with the degree of uniqueness, comparing the original untranslated pattern with the set of translated patterns. Uniqueness could also be characterized in contrast with other reference patterns from different peptides or peptide fragments (each appropriately translated to the parent m/z) or other ways of scrambling the target reference pattern different from m/z translation.

The probability as a function of m/z translation and the event number M forms a surface $P(M, m/z \text{ offset})$. We typically represent this surface by a single probability contour at some constant P_0 , i.e., plotting M such that $P(M, m/z \text{ offset}) = P_0$ as a function of the m/z offset (Figure 1). Detection confidence along the fixed probability contour is associated with the M amplitude of the 0-offset peak, $M_{P_0}(0)$, relative to the M amplitude at other m/z offsets, $M_{P_0}(\tau)$.

METHODS AND EXPERIMENTAL DETAILS

Detection Confidence Score. Detection confidence in a single $M_{P_0}(\tau)$ plot is related to ratios between the number of events that can be associated with the reference distribution in its original position, i.e., $M_{P_0}(0)$, and the number of events that can be associated with the reference distribution when it is translated by τ , $M_{P_0}(\tau)$. A lack of relative amplitude is clear nondetection as the defining events for the reference pattern are not above the ion background. An $M_{P_0}(0)$ peak above

background, however, invites the question of its statistical significance and a common approach for characterizing detection significance is to define a score reflecting some function of the relative amplitudes in the $M_{P_0}(\tau)$ plot and use Receiver Operating Characteristic (ROC) analyses¹² to identify the trade-off between this score and detection accuracy. A reasonable score parameter is the ratio of the “0-offset” peak $M_{P_0}(0)$ to some measure of the fluctuation in the score as the reference pattern is translated. Specifically, let k be a set of discrete ± 1 amu translations of the reference pattern, e.g., discrete translations from -50 to $+50$ amu, calculate an averaged event number $\bar{M}_{P_0} = (1/101) \sum_{k=-50}^{50} M_{P_0}(k)$ and subtract this from the 0-offset peak. This difference is then normalized by the averaged fluctuation (eq 4).

$$s = \frac{(M_{P_0}(0) - \bar{M}_{P_0})}{\sqrt{\frac{1}{101} \sum_{k \in S_M} (M_{P_0}(k) - \bar{M}_{P_0})^2}} \quad (4)$$

In contrast to LC-MS/MS analyses, MS³ experiments for low level detection generates only a few spectra so large experimental MS³ data sets for ROC analyses are more difficult to establish than with typical data-dependent LC-MS/MS proteomics analyses. Moreover, as detection scores reflect both the target and ion background, the utility of a highly averaged performance parameter for detecting a specific target is limited.

Multiple MS³ Spectra. Another approach to improving confidence is to examine multiple MS³ spectra for a single target. The first stage of dissociation of MHC I bound peptides generally produces multiple MS² fragments that can be used for MS³ detection. All of the expected MS³ patterns must be detected for the target to be present. However, one must be careful about what is expected. Contradictory detection, a situation in which one fragment is identified with a high score but a second fragment is not, could reflect a false positive in

(11) Yates, J. R., 3rd; Morgan, S. F.; Gatlin, C. L.; Griffin, P. R.; Eng, J. K. *Anal. Chem.* **1998**, *70*, 3557–3565.

(12) Fawcett, T. *Pattern Recognit. Lett.* **2006**, *27*, 861–864.

the first detection, but it could also arise from inadequate ion signal in the second, i.e., if the ion flux distribution identifying the second fragment does not rise above the Poisson background. For example, an MS² fragment might be of low abundance, and/or the ion background could be high, and/or the MS³ dissociation of the fragment could be dominated by neutral losses, which makes for a poor detection signature. As multiple MS³ spectra are collected in an alternating sequence to correct for the variations in time of the nanospray source, the relative ion counts between the two MS³ spectra, like the relative intensities of peaks within a single MS³ spectrum, is a measurable property of the target peptide under fixed instrumental conditions. If one determines the level of one MS³ reference pattern in a sample containing the target, then one knows the levels at which other MS³ reference patterns of the target should be observed. For the reference (e.g., a synthetic peptide), the relative ion counts between different MS³ spectra is simply the ratio of summed events in each spectrum. For a complex mixture, background fragment peaks will overlap and obscure the events in the mixture spectrum that can be definitely assigned to the target, but this is essentially what the Poisson 0-offset amplitude measures. As this magnitude depends in some degree on the cutoff probability P_0 and the set of peaks included in the reference pattern, the relative event counts for both the synthetic peptide and the target in the mixture are consistently determined by Poisson fitting with common reference files and cutoff probability P_0 . These event ratios can be used to discriminate between a false positive and inadequate ion signal (see Supporting Information for further details).

Quantitation. Two characteristics of the general MS³ analyses are worth emphasizing. First, different MS³ spectra are acquired using alternating scans so that the relative ion fluxes among different MS³ spectra are measured. Second, since the mixture is not separated, quantitation reference molecules are present in the ion beam at all times. There is no need to add stable isotope analogues that coelute to determine the ion flux at the time of the target's elution. One does assume that for fixed relative concentration in a mixture the relative ion events arising from a target and a calibrant molecule are also constant. This is not just a matter of ionization suppression from solutions that are too concentrated. For some peptides in dilute and simple mixtures, the relative ion flux depends significantly on handling and surface exposure, but when the target and calibrant are minor components of a complex mixture of endogenous MHC I peptides, a constant relative flux is more likely to be observed. Although we do not fully understand the physical and chemical processes behind the carrier role played by the large excess, it seems that a part of this is due to a normalization of spray conditions and a blocking of surface adsorption sites. Quantitation is generally based on the ratio between ion counts assigned to the target and ion counts that track the sample load, i.e., the product of ion flux and collection period. Poisson assignment is used to extract the target and load-specific ion counts from the MS³ spectra, addressing the high probability of peak overlap with chemical background.

In some analyses, one can first add a calibrated amount of target peptide to a negative control and determine the ratio of

ion counts between background and target ion fragments (see Results, below). If the experimental system does not provide a good negative control (e.g., the target is a peptide presented by a cancer cell line), then one can quantitate relative to an added calibrant. Here, a known amount of both target and calibrant are added to a carrier system (without target or calibrant molecules) and MS³ ion fluxes are measured. When the sample is analyzed, only calibrant is added to the sample and quantitation is relative to the calibrant.

Cell Lines. For detection of viral epitopes, BEAS-2B cells,¹³ transfected to express surface HLA-A*0201, were infected for 18 h with 10:1 multiplicity of infection (MOI) of influenza PR8 or left as uninfected controls. For quantitation studies, the T1 cell line (HLA-A2+) from ATCC (CRL-1991) was used. Cells were grown in Dulbecco's modified eagle medium (DMEM; Sigma, St. Louis, MO) supplemented with 10% fetal calf serum (FCS), 1% L-glutamine, and 1% penicillin/streptomycin.

Nanoscale HLA-A*0201 Purification by Immunoprecipitation. For each immunoprecipitation, 10 μ g of anti-HLA-A02 BB7.2 mAb (BD Biosciences) was noncovalently coupled to 20 μ L of Gamma Bind beads (GE Biosciences) for 1 h at room temperature. BEAS-2B or CaSki cells (10^7) were harvested during the log growth phase and washed with PBS. Cells were pelleted, and the washed and dried cell pellet was lysed using 1.5 mL of lysis buffer consisting of 20 mM Tris, pH 8.0, 1 mM EDTA, and 100 mM NaCl; 1% Triton X100; 60 mM *n*-octylglucoside; protease inhibitor tablet (Roche Biosciences); and 1 mM PMSF for 10 min on ice. Cell debris was removed using centrifugation for 30 min at maximum speed (13000 rpm) at 4 °C. Cleared supernatant was incubated with 20 μ L of antibody coupled Gamma Bind Plus beads for 2 to 3 h at 4 °C. Beads were washed four times using lysis buffer without Triton-X-100 and protease inhibitors. Beads were further washed four more times with 10 mM Tris, pH 8.0. All free buffer was aspirated, and the wet antibody-HLA beads were immediately stored at -80 °C.

Acid Extraction and Peptide Isolation. Peptides were eluted from the beads with 17.5 μ L of 10% acetic acid at 65 °C for 15 min. Antibody, peptide, MHC heavy chain, β_2m , and nonspecifically bound polypeptides are trapped on a 1 μ L C18 reverse phase tip. The antibody is not covalently bound to the beads, and there is no filtration of the eluate. The tip is washed with 0.1% TFA, followed by 0.1% formic acid, and then 20% MeOH and eluted with 2 μ L of 60% MeOH in 0.1% formic acid.

MS and MS³. All MS data were generated by static nanospray. An AB Sciex QTrap 4000 quadrupole-LIT mass spectrometer was used for MS³ analysis, and an AB QSTAR Elite Q-TOF was used for MS analysis. MS³ spectra are always reported by the data system as discrete events.

RESULTS

Detection of Viral Peptides from Influenza A Infection of Human Bronchial Epithelial Airway (BEAS-2B) Cells. HLA-A*0201 peptides were isolated from infected [10:1 MOI, A/PR8/

- (13) Reddel, R. R.; Ke, Y.; Gerwin, B. I.; McMenamin, M. G.; Lechner, J. F.; Su, R. T.; Brash, D. E.; Park, J.-B.; Rhim, J. S.; Harris, C. C. *Cancer Res.* **1988**, *48*, 1904-1909.
- (14) Chernushevich, I. V.; Bahr, U.; Karas, M. *Rapid Commun. Mass Spectrom.* **2004**, *18*, 2479-2485.

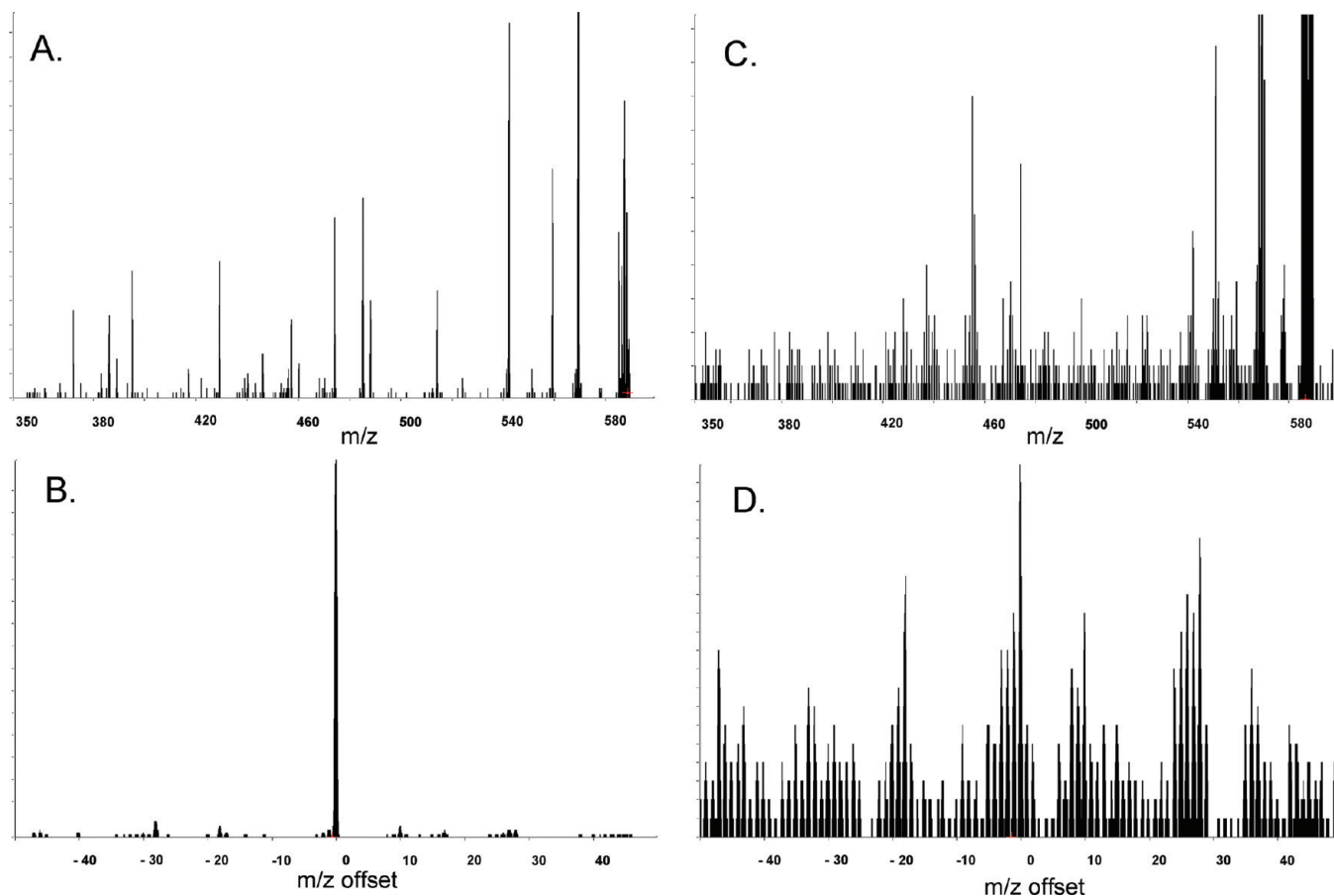


Figure 2. MS³ detection of immunodominant M1_{58–66} influenza peptide GILGFVFTL from 10⁷ infected and uninfected human bronchial epithelial (BEAS) cells. (A) 90 s collection of MS³ 483.8:587.4 of HLA–A2 peptides isolated from infected BEAS cells produces a spectrum almost indistinguishable from the synthetic peptide (synthetic spectrum not shown). (B) Poisson detection plot for the *M*-dependent probability of the reference b₆ (GILGFV-) fragment's dissociation pattern in MS³ spectrum of panel A. (C) 9 min collection of MS³ 483.8:587.4 from uninfected BEAS cells. (D) Poisson detection plot for the *M*-dependent probability of b₆ fragment in MS³ spectrum C. In all four plots, the *y*-axis is in units of “events”; for A and C, it is recorded events; for B and D, it is *M* of Figure 1.

34(H1N1)] (Charles River) and uninfected samples of 10 million BEAS-2B cells. A Q-TOF MS nanospray spectrum of the unfractionated peptides recovered from infected cells shows the very characteristic signature of doubly charged MHC I peptides (see Supporting Information, Figure S1). Neither singly nor triply charged peptides are observed with abundance, although low levels of peptides in these charge states can be identified by characteristic amino acid residue losses in MS/MS spectra. Combining the known proteome for the influenza A strain (PR8) with a ranking based on *in silico* binding predictions (<http://www.immuneepitope.org/>), a set of influenza peptides was identified as targets for MS³ Poisson detection on infected HLA-A*0201⁺ cells. The time course and range of influenza peptide presentation in bronchial epithelium and other cell lines will be discussed in a future manuscript. Here, the detection method is the focus, and illustrative data will be shown for the MS³ Poisson detection of two peptides in the target list, one very abundant and the other present at low levels. The influenza M1_{58–66} peptide is a well-known immunodominant T cell epitope, and it can be immediately identified in the MS³ 483.8:587.4 spectrum of the unfractionated peptides after a 90 s collection (Figure 2). There is no need for probabilistic identification; the MS³ spectrum from the sample is almost indistinguishable from that of the synthetic peptide (synthetic peptide data

not shown). After a 31 min collection, the PB1 peptide FVANFSMEL was also detected and detected only in the infected BEAS-2B cells (Figure 3). Here, the MS³ 529.3:579.3 spectra from infected and uninfected cells are quite similar in appearance, and it would be difficult to identify by visual inspection that the fragmentation pattern of the b₅ ion FVANF⁺ is present in the infected but not in the uninfected spectrum. The probabilistic metric makes this identification handy (Figure 3C,D).

Quantitation of Surface Copies in Loading T1 Cells with Influenza Peptide GILGFVFTL. An important assay in immunology is to load MHC I or II molecules on the surface of antigen presenting cells (APCs) with a peptide; add T cells; and measure cytotoxicity, cytokine production, or other markers of T cell recognition of the MHC/peptide complex on the surface of the APC. Of particular interest is a molecular characterization of the avidity or quality of the T cells in these assays, specifically, the number of peptide copies per APC required for T cell recognition. By incubating APCs with increasingly dilute peptide solutions and measuring the T cell response when they are placed in contact with these cells, one can prepare APCs that are loaded with a minimal amount of peptide for T cell recognition. The immunological context and experimental details are being prepared for a future publication. Here, we will focus on the analytical

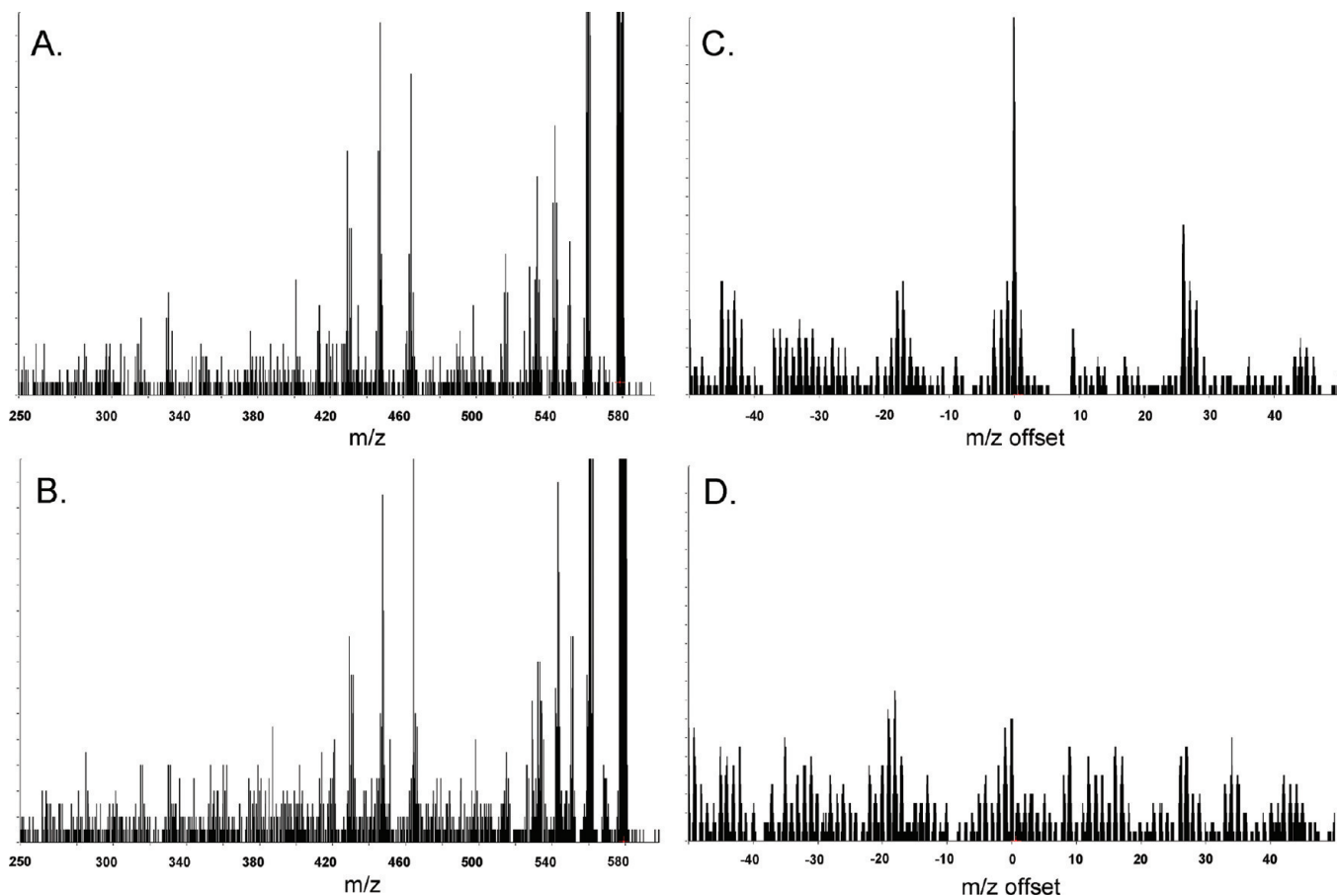


Figure 3. MS³ Poisson detection of influenza peptide FVANFSMEL in 10 million infected and uninfected human bronchial epithelial cells. (A) MS³ 529.3:579.3 spectrum from infected sample. (B) MS³ 529.3:579.3 from uninfected sample. (C) Poisson detection plot for the M -dependent probability of the reference b_5 (FVANF-) fragment's dissociation pattern in spectrum A. (D) Poisson detection plot for the probability of reference b_5 pattern in spectrum B. The magnitude of the 0 m/z offset peak relative to nonzero m/z offsets is the metric of detection. As in Figure 2, the y -axis units are "events".

method: how Poisson MS³ quantitation is used to determine the copies per cell of minimally loaded APCs.

T1 cells expressing high levels of HLA-A2 were used as the APC and the immunodominant peptide GILGFVFTL from the influenza M1 protein served as the cognate peptide. The T1 cells were loaded by placing them in solutions of the peptide covering a range of concentrations, and a T cell line recognizing this peptide determined a minimal loading of the T1 cells at 62 pg/mL by the INF- γ ELISPOT assay. Three samples were prepared: 5 million unloaded T1 cells, 5 million T1 cells with 829 amols (100 peptide copies/cell) of GILGFVFTL added to the affinity beads containing the HLA-A2 complexes at the start of the acid elution step, and 5 million T1 cells loaded with GILGFVFTL peptide at a concentration of 62 pg/mL prior to affinity purification of HLA-A2 complexes. For each sample, the MS³ spectra 483.8:587.4 (targeting the GILGFVFTL b_6 fragment) and 483.8:796.5 (targeting the y_7 fragment) were acquired in alternating sequence. Poisson plots for detecting GILGFVFTL were calculated (Figure 4). To quantitate, one wants to determine the GILGFVFTL-specific ion counts relative to T1 background counts. The unloaded T1 cell sample produced a weak ion background in the MS³ 483.8:587.4 spectrum, so just the MS³ 483.8:796.5 spectrum was used as a reference pattern for the T1 background. The Poisson fits (0-offset amplitude) of MS³ 483.8:796.5 spectra to the reference T1 background for each sample was calculated to measure the

product of ion flux (from endogenous HLA-A2 bound peptides) and collection period. The 0-offsets measuring GILGFVFTL-specific ion counts (Figure 4) were then normalized for each sample by dividing by this product. The T1 background contributions to the 0-offsets for the GILGFVFTL Poisson fits were also calculated. The corrected 0-offset amplitudes of Poisson fits to the three reference patterns (T1 background and GILGFVFTL y_7 and b_6) for all three samples are summarized in Table 1 (see Supporting Information for further details).

Two partially independent estimates of GILGFVFTL amounts in the "T1 + 62 pg/mL" sample can be made, one from the y_7 and one from the b_6 fragment. For the y_7 fragment, 99 events in the "T1 + 62 pg/mL" sample, scaled (divided) by the 1855 T1 background events, is 9.8% of the 314 events in the "T1 + 829 amols" sample scaled by the 577 T1 background. For the b_6 fragment, 166 divided by 1855 ("T1 + 62 pg/mL" sample) is also 9.8% of 528 divided by 577 ("T1 + 829 amols" sample). As the "T1 + 829 amols" sample added peptide calibrated to 100 copies per cell, both the y_7 and b_6 MS³ spectra indicate 9.8 copies of GILGFVFTL per cell is loaded on T1 cells by incubation in a solution of peptide at 62 pg/mL.

If one attempts quantitation using a metric algorithm (e.g., the correlation function), the interference of the chemical background would prevent the assignment or even the detection of the GILGFVFTL peptide in T1 cells loaded at 62 pg/mL (Figure 5).

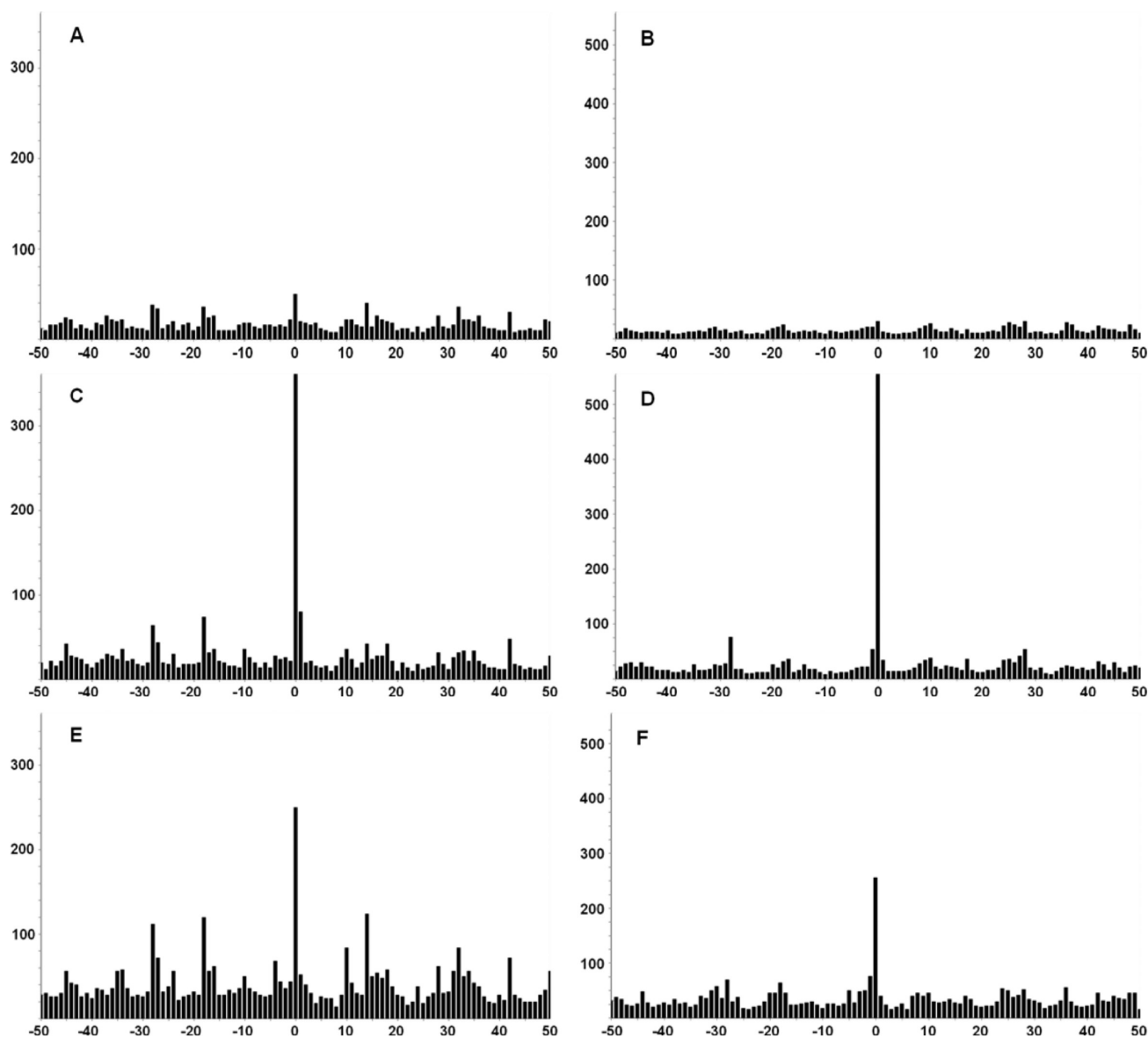


Figure 4. Poisson detection plots for peptide GILGFVFTL in samples of 5 million T1 cells. (A, C, E) Detection of y_7 fragment in the MS³ spectrum 483.79/796.46; (B, D, F) detection of b_6 fragment in the MS³ spectrum 483.79:587.34. The x-axis is m/z translation of the reference pattern; the y-axis is the event number corresponding to the cutoff probability. (A, B) Five million T1 cells, the negative control. (C, D) Five million T1 cells with 829 amols or 100 copies/cell of GILGFVFTL added. (E, F) Five million T1 cells incubated in a solution of 62 pg/mL GILGFVFTL prior to cell lysis and affinity purification of HLA–A2 complexes.

Table 1

	T1 background	y_7 MS ³ 483:796	b_6 MS ³ 483:587
T1 blank	616	50	30
T1 + 829 amols	577	314	528
T1 + 62 pg/mL	1855	99	166

DISCUSSION

The nanospray MS³ Poisson detection method characterized here measures and analyzes MS³ spectra of a complex sample using a hybrid quadrupole-LIT instrument. The detection performance of this, relative to other MS methodology, is related to aspects of molecular separations, instrument duty cycle, and signal recognition, topics that are broadly involved in MS analyses.

Conventional MRM detection fixes (more generally schedules) the m/z window for parent ion selection by a mass filter, fragments the selected window, and monitors previously identified fragment windows with a downstream mass filter. Again molecular separations must be used, for is it a chromatogram of (some function of) the ion abundances in the monitored windows that is the detection signature. A significant aspect of the MS³ Poisson detection sensitivity is avoiding separations. In identifying limits of detection, an absolutely fundamental parameter is the conversion of molecules in the condensed phase into gas phase ions. With electrospray ionization, it is generally well-known that lower flow rates improve sensitivity. This reflects both a more efficient ionization (smaller droplets, less clustering, higher charge to analyte ratio) and that a higher fraction of the ion plume is transferred through the conduction limit into the operating vacuum of the instrument. In reducing nanoLC

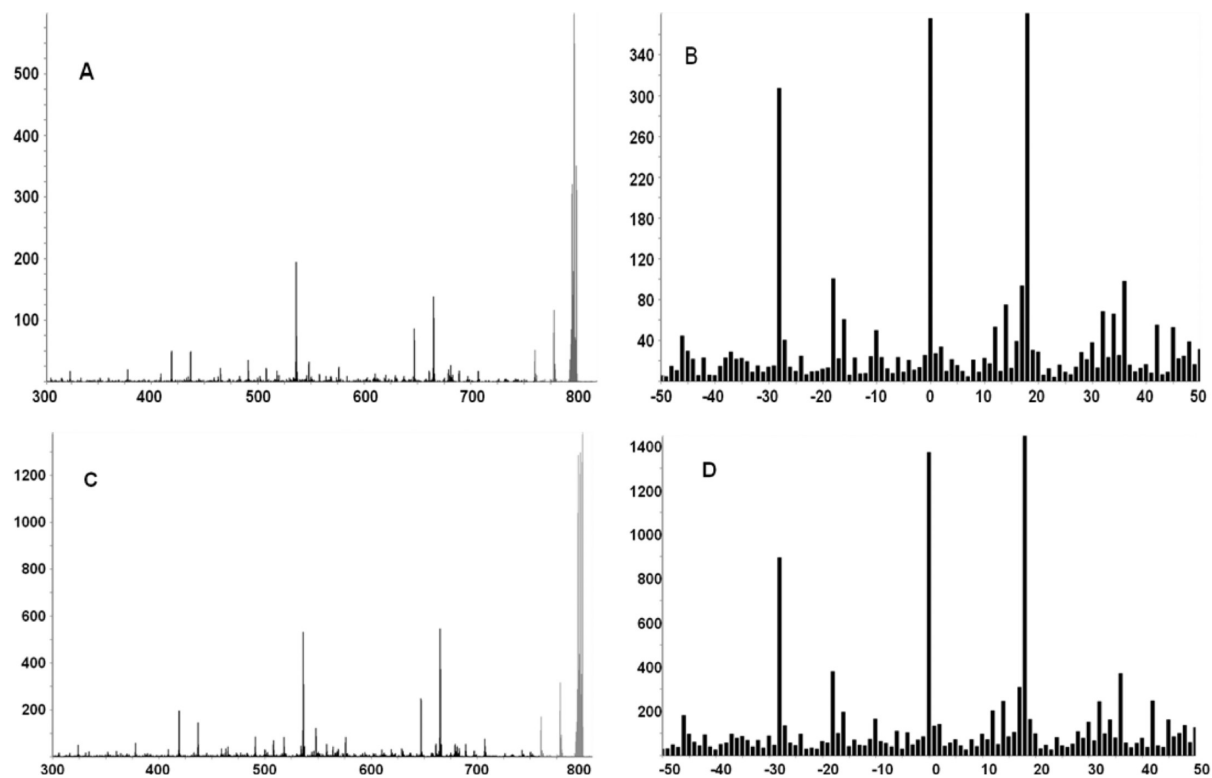


Figure 5. MS³ 483.8:796.5 spectra of control T1 cells (A) and 62 pg/mL loaded T1 cells (C) have only minor spectral differences due to the dissociation of the γ_7 fragment of GILGFVFTL. The overlap of major background fragments with some of the peaks expected in γ_7 's dissociation dominates the metric contrast and its associated sliding inner product or correlation function score (B, D). The negative sample (A, B) cannot be distinguished from the positive pair (C, D). This is in contrast to the Poisson plots of Figure 4A,E made from the same reference pattern and MS³ spectra. To improve the detection specificity of the translated inner products (B, D), the m/z range above 760 is removed (shown in gray) as the peaks in the parent window near m/z 796.5 and the associated neutral losses have little value in detection but add substantially to the 0-offset amplitude and at offsets corresponding to neutral mass increments in the correlation function plots.

flow rates to improve ionization, column and capillary dimensions must also decrease to avoid diffusional broadening, and this increases surface to volume ratios and provides additional opportunity for surface adsorption. Even in direct loading of the tip of the nanospray needle, adsorption is a nontrivial issue; a problem that is not uniform among different analytes, and a problem that is not just restricted to borosilicate glass needles, although the capacity of borosilicate to adsorb certain peptides is impressive.¹⁴

A general objection to nanospray analyses of complex mixtures without separations due to "ion suppression" has often surfaced in past discussions, and this objection merits some attention. To first order, electrospray is a constant current source, and "suppression" of charged solvent cluster ions by surface active, higher proton affinity peptides is also a reason for the sensitivity of electrospray. Suppression by surface active ionic detergents, on the other hand, is a well-known problem. The objection that has been phrased is not so much suppression in general, but the suppression of analytes by other analytes. To be concrete, consider ion suppression in the analysis by static nanospray of a 1 μ L sample of MHC I peptides. Given that nanospray's smaller droplets and higher charge to analyte ratio decreases suppression, a concentration limit of 10^{-4} to 10^{-5} M before peptide/peptide suppression dominates is consistent with literature results.¹⁵ If each cell has 10^5 peptide-MHC complexes (a high level)

and one recovers through affinity purification, acid extraction, and C18 trapping 100% of the peptides from 10 million cells and puts all these peptides in a 1 μ L volume, the resulting concentration is 1.66×10^{-6} M. Ionization suppression of analyte by analyte certainly arises in principle but is a phenomenon characterized by well understood material scales.

For an ion trap operating in MS/MS mode, the trap is first filled to its charge capacity and then the parent ion is isolated and dissociated, and the fragments are scanned out. If the incident ion flux is distributed in m/z , only a small fraction of the trapped ion population is in the selected m/z window, resulting in a low number of target ions for each MS/MS cycle of the LIT. If the fill time is a small fraction of the overall MS/MS cycle time (i.e., a low duty cycle), the potential sensitivity that could be achieved is correspondingly reduced. There are a number of ways to selectively accumulate ions in an LIT (RF level/DC offset between rods, resonant ejection during accumulation, injection waveforms), but these involve various compromises in trapping efficiency, losses due to off-resonant excitation, and selection resolution. Quad-TOF instruments do not suffer the same degree of performance degradation with increased ion beam complexity. Here, the duty cycle is not coupled to the source ion flux and m/z distribution but to the fraction of the mass-filtered and fragmented ion beam that is extracted for TOF analysis. This fraction is generally small but with caveats concerning mass range and extraction methods. The MS³ molecular detection experiments described in this study used a hybrid geometry in which a

(15) Tang, K. Q.; Page, J. S.; Smith, R. D. *J. Am. Soc. Mass Spectrom.* **2004**, *15*, 1416-1423.

quadrupole mass filter transmits the parent ions, which are subsequently fragmented in a collision cell, and the fragments were collected in a downstream LIT (QTrap 4000). Similar to quad-TOF, the target molecular ion isolation and the first fragmentation are operated in a beam mode. The accumulation charge capacity of the downstream LIT is devoted only to the charge transmitted in the selected m/z window. In addition, the isolation of an MS^2 fragment for MS^3 is by DC offset applied to the rods, which is more stable to charge loading than resonant excitation. Although the selection resolution by DC offset is low (3 to 4 amu), the resolution of the resonant excitation for dissociation is high (<1 amu), and since the excitation is applied only after the second isolation, the charge load in the LIT at this point in the MS^3 cycle is low. As a result, the MS^3 cycle in our analyses is dominated by long collection periods (>500 ms) and a correspondingly high duty cycle.

Following ionization and instrument duty cycle, signal recognition is the third component of the MS^3 detection method. Our approach to detection has been theoretically direct: characterize the underlying random process and identify the probabilistic "distance" that naturally arises in sampling this process. Although we agree with the inherent pragmatism in the observation that a good scoring system does not necessarily have any theoretical foundation,¹⁶ one ignores foundations at some peril. The use of a metric distance (least-squares, spectral angle, sliding inner product, correlation function) for comparing spectra and scoring the result is especially ineffective when the spectra contains substantial and structured chemical noise. "Noise" is a poor term for what is often other related molecules, e.g., coeluting or coisolated, since there is an association of a statistical regularity with spectral noise that does not at all apply here. If one knows the target's spectrum (here an MS^3 spectrum), then treating MS measurement as sampling a random process improves discrimination. A metric approach generally fails because it overweighs the significance of m/z coincidence in the big peaks and gives too little negative weight to circumstances in which minor peaks are not observed. Comparing the correlation function or sliding inner product of Figure 5B,D with the Poisson plot of Figure 4A,E, calculated for the same y_7 reference pattern and MS^3 483.8:796.5 spectra, illustrates the problem of metric contrast. Here, a few large background peaks in both measured spectra line up with reference peaks and dominate the inner product. The expected target peaks are all small relative to these background peaks. That they are not present has little metric consequence given the overlap of the large peaks. From the standpoint of sampling probability, expecting 20 events in a data channel and observing none is high confidence elimination, irrespective of how many expected events were observed in another data channel. At a fundamental level, the Euclidean or least-squares difference (in all of its implementations) between a reference pattern and observed data is simply not a good parameter for discrimination of an arrival rate process (see Supporting Information for further details).

Conversely, where peptide identification from MS and MS/MS spectra involves generating in silico model MS/MS spectra from peptide sequences and then matching the model spectra to

the observed spectra, probabilistic detection can fail where the permissive metric detection will not. If the model calculates an ion of even minor relative amplitude that is not produced in the physical dissociation, the impact on probabilistic detection could be severe whereas the fractional loss of ion current in an inner product matching would be minor.

A connection between the asymptotic probability, relative entropy, and a distance between a model and an observed spectrum is not new.¹⁷ What is new is implementing the connection when the model spectrum is embedded in a chemical background of coselected ions or ion fragments. In many settings, an m/z window selected for dissociation has multiple coeluting ion species and a direct relative entropy distance between a library spectrum and a data spectrum with chemical noise present can fail to detect¹⁸ because overlapping chemical noise can result in a high entropy cost. In evaluating different scoring algorithms for spectral matching, the degree of chemical noise is a very important parameter; important because it is the lower intensity spectra that challenges existing matching algorithms, and it is the lower intensity spectra that are more likely to suffer from significant chemical noise. There is wide recognition of the utility in archiving MS/MS spectra of known peptides and using a library of archived spectra for reference patterns. Mixtures of proteome scale complexity, incompletely resolved by a given stage of separations, lead to MS/MS spectra with fragment overlap. As in MS^3 detection, the basic need is to detect in the MS/MS spectra known event limited patterns against nontrivial ion backgrounds, and for this, an $L^p(N)$ metric distance may be less effective than detection based on the Poisson methodology described above.

For targeted detection, we have examined formally the probability of event patterns in the context of a random process with a substantial ion background and finite event sampling. This probability, plotted against a parameter (m/z displacement) to characterize uniqueness, serves as a detection metric. The metric allows high specificity in the identification of a target component in a complex dissociation spectrum with multiple components that overlap with target fragments. However, the dissociation pattern of the target must be known with high confidence. The specificity, combined with two stages of ion selection (parent ion and MS/MS fragment ion), allows the detection analysis of complex mixtures by nanospray MS^3 without molecular separations. Low flow static nanospray improves ionization efficiency, and in restricting analyte adsorption by restricting surface exposure, allows a decrease in the material scale of the sample being analyzed. That the mixture is not separated also aids in target quantitation. Without separations, the ion beam contains a complex population of molecular species. From the standpoint of ion devices, a complex ion stream is difficult to analyze in the fill, isolate, dissociate, and scan sequence of a single LIT due to a low duty cycle. Selectively filling the LIT with a quadrupole mass filter produces a much higher duty cycle, and this leads to better operational sensitivity. The meth-

(17) Reinhold, B. B.; Reinhold, V. N. *J. Am. Soc. Mass Spectrom.* **1992**, *3*, 207–215.

(18) Zhang, Z.; Sun, S.; Zhu, X.; Chang, S.; Liu, X.; Yu, C.; Bu, D.; Chen, R. *BMC Bioinf.* **2006**, *7*, 222.

(16) Fenyo, D.; Beavis, R. C. *Anal. Chem.* **2003**, *75*, 768–774.

odology can be applied to the detection of immunologically significant peptides associated with MHC I molecules that have been targeted by genetic, immunologic, and/or MS discovery analyses. Physical detection at high sensitivity or low copy number, accurate quantitation, and the ability to analyze small numbers of cells all allow for understanding the presentation of defined tumor- and infectious disease organism-derived epitopes on relevant cell populations. This will aid in aspects of diagnosis (biopsy or other tissue) and assist with the rational design of vaccines based on T cell mediated immunity.

ACKNOWLEDGMENT

This work was supported in part by U19 AI57330 from NIAID and DFCI Cancer Vaccine Center funding.

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 7, 2010. Accepted September 25, 2010.

AC102387T