High Resolution Mass Spectrometry for Rapid Characterization of Combinatorial Peptide Libraries

Magnus Palmblad,*,† Jan W. Drijfhout,‡ and André M. Deelder†

Biomolecular Mass Spectrometry Unit, Department of Parasitology, and Department of Immunohematology and Blood Transfusion, Leiden University Medical Center, Leiden, The Netherlands

Received August 13, 2009

Combinatorial libraries can be characterized in detail by high resolving power mass spectrometry. We here demonstrate this for synthetic decapeptide libraries using a state-of-the-art 15 T Fourier transform ion cyclotron resonance mass spectrometer. Using comparison between predicted and measured spectra, a number of metrics can be derived that shed light on the library composition and degeneracy of elemental compositions. These techniques can be used to rapidly quality control combinatorial synthesis products or follow combinatorial libraries in more detail, for instance during binding studies or chemical reactions. The methods are in principle equally applicable to the analysis of other types of combinatorial libraries.

Introduction

Combinatorial synthetic chemistry has been used to create molecular libraries ranging from hundreds to trillions ($>10^{12}$) of compounds.[1] Combinatorial libraries can be constructed from peptides[2−6] and nucleotides, but also from a wide range of other building blocks.[7] Combinatorial peptide libraries have a diverse range of applications and have been used for studies of antigen−antibody interaction on the amino acid level,[2,5,8] combinatorial drug discovery,[9] chemosensors,[10] and finding inhibitors of enzymes[11] or aggregation in proteinopathies.[12] More recently, hexapeptide libraries have been used to enrich low-abundant proteins in proteomics.[13] Peptides can also be used for encoding compounds (beads) in other one-bead-one-compound libraries.[14]

Mass spectrometry is a powerful tool to monitor combinatorial syntheses, estimate yields, and follow screening experiments. Matrix-assisted laser desorption ionization (MALDI), often with time-of-flight (TOF) mass spectrometry, is a common method for characterizing or screening combinatorial libraries, including peptide libraries.[15,16] Liquid chromatography or ion mobility techniques have been used in combination with mass spectrometry to analyze larger peptide libraries (>1,000 compounds)[17,18] However, even such highly complex mixtures can be analyzed directly by high resolving power mass spectrometers, such as modern TOF, Orbitrap,[19] or Fourier transform ion cyclotron resonance (FTICR)[20] instruments. The use of FTICR in the analysis of large combinatorial peptide libraries was discussed early on by Demirev and Zubarev[21] and demonstrated by Nawrocki et al. using 4.7 and 9.4 T instruments.[22] The main advantage of direct analysis is speed: a single mass spectrum can be acquired in a matter of seconds rather than the tens of minutes required if mass spectrometry is used in hyphenation with a liquid separation. High magnetic field FTICR instruments provide the highest resolving power of commercially available mass spectrometers and are capable of resolving tens of thousands of species of similar abundance in a single spectrum.[23,24] For direct analysis of complex mixtures, dynamic range is also critical. The dynamic range is ultimately limited by the number of ions that can be trapped in the FTICR cell and measured simultaneously. This number, along with other performance characteristics, increases quadratically with the magnetic field,[20] which explains the interest in and push for higher field instruments. The current state-of-the-art in FTICR mass spectrometry is represented by a small number of 15 T instruments installed in laboratories across the world. In this paper we want to revisit the application of FTICR mass spectrometry for the direct analysis of large combinatorial peptide libraries using a high-field instrument and discuss metrics for comparison of calculated and measured high-resolution mass spectra.

Results and Discussion

In general, we found the agreement between predicted and measured library spectra to be remarkably good for the tested peptide libraries (Figures 1 and 2). Not surprisingly, shorter, lighter, non-predicted peptides could be observed. These are likely due to incomplete synthesis. The total abundance of such truncated peptides was estimated at <1% of the total yield. No significant signal corresponding to longer peptides was observed. Mass accuracy is paramount when analyzing complex samples, comparing a large number (10^3-10^5) of measured masses with an equally large or larger number of possible, calculated masses. For instance, the calculated spectrum for the complete $X_1X_2X_3$GLYNLVK library contains 22,120 peaks at $\geq 10^{-3}$ abundance relative to the base peak with a bin size of 1 mDa. The 15 T FTICR mass spectrometer used to analyze the peptide libraries routinely achieves <200 ppm mass measurement standard errors for simple samples, when all species can be well resolved (see

* To whom correspondence should be addressed. E-mail: n.m.palmblad@lumc.nl.
† Biomolecular Mass Spectrometry Unit, Department of Parasitology.
‡ Department of Immunohematology and Blood Transfusion.

10.1021/cc9001235 CCC: $40.75
Published on Web 10/28/2009

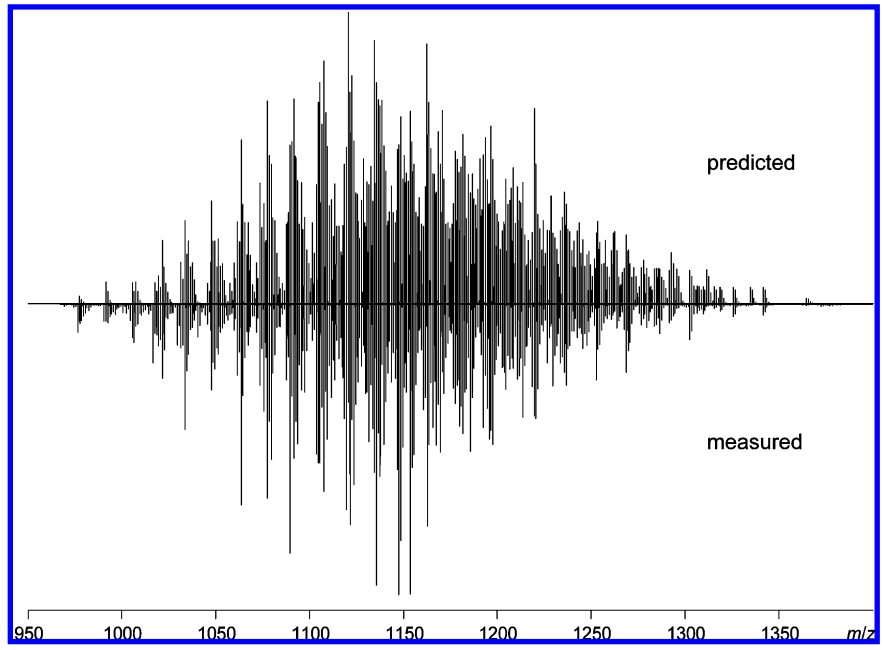

Figure 1. Predicted versus measured mass spectra for a combinatorial peptide library $X_1X_2X_3$GLYNLVK where X_i is any of the 20 naturally occurring amino acids. The library was analyzed by direct infusion electrospray ionization on a 15 T FTICR mass spectrometer. The spectrum shows the full measured range of singly charged ions. A similar distribution of doubly charged species was observed in the *m*/*z* 480−670 range.

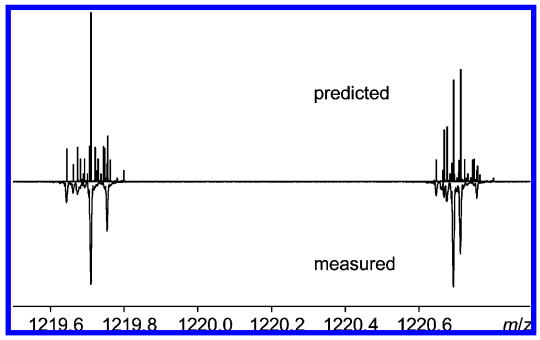

Figure 2. Small part of the singly charged region of the full spectrum in Figure 1, illustrating that detailed information can be obtained from very complex samples with high resolution mass spectrometers. Many (though not all) predicted species are also observed, and measured intensities generally correlate well with predicted intensities. A small amount of shorter (and therefore lighter) peptides was also observed. A bin size of 0.001 Da and uniform assumed ionization and detection efficiency were used in the prediction of the library spectrum.

Figure 3a). This is not the case with extremely complex samples (Figure 3b), where the mass measurements of some species are affected by partially overlapping peaks. Although it is theoretically possible to compensate for peak overlaps to some degree, this is not a trivial task for unknown species and unknown numbers of components. The high resolving power in comparison with calculated isotopic fine structure allows more detailed analysis of combinatorial libraries than would be possible with lower resolution instruments, e.g., a MALDI-TOF in linear mode. Comparison, for instance cross correlation, between the predicted and measured library mass spectra is a direct quality control metric on the synthesis product. The correlation will never be perfect, but deviations from the normal level of agreement between the spectra are easy to detect. Large errors in the synthesis producing many smaller or larger compounds will be detectable with almost any type of mass spectrometry, but high resolution mass spectrometry allows detection also of minor errors such as incorporation of wrong building blocks or poor reaction yields. Finally, we could observe that *on average*, the signal intensity at a given *m*/*z* or unique elemental composition is proportional to the degeneracy of that elemental composition in the combinatorial peptide library (Figure 4). This analysis thus reflects the distribution of elemental compositions of the entire 8,000-peptide combinatorial library.

Conclusion

We have shown how high resolution mass spectrometry can be used to characterize complex combinatorial peptide libraries by simple and rapid direct infusion analysis. The strategy of comparing calculated theoretical and measured spectra for entire libraries is not limited to peptides but can be applied in the mass spectrometric analysis of any type of combinatorial library. The relative frequency of degenerate elemental compositions in the library compare well with the measured average signal intensity. We speculate that with the increased availability of high resolving power instruments, the rapid, high-throughput method described here can be used to monitor combinatorial library compositions during binding or chemical reactions in real time.

Experimental Section

Synthesis of Combinatorial Peptide Libraries. Twenty separate sublibraries of the form NH_3-$X_1X_2X_3$GLYNLVK-COOH, where X_1-X_3 were varied over all 20 naturally occurring amino acids, where synthesized from the C- to the N-terminus using Tentagel S AC as a resin.[8] In short, Fmoc-based solid phase peptide synthesis was performed in 20 separate reactors (10 μmol resin loading each). After coupling cycle 8 and 9, the resin beads of all reactors were mixed and re-divided over the reactors.[3] The 20 libraries with different X_1 were kept separate after the last step.

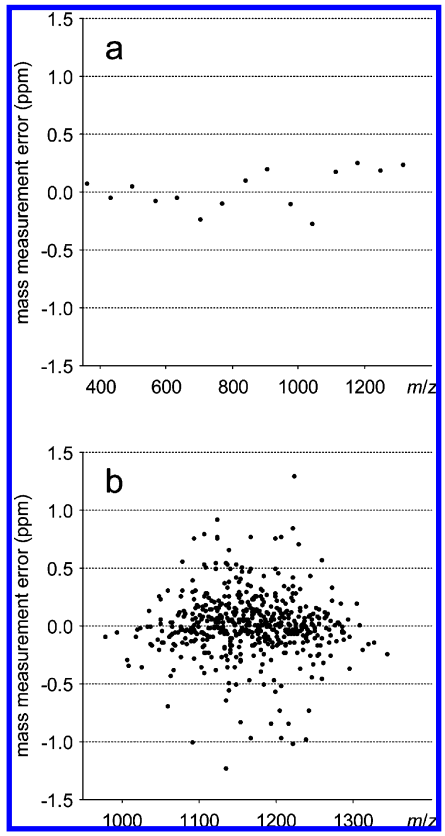

Figure 3. 15 T FTICR mass spectrometers routinely provide sub-parts-per-million mass measurement uncertainty with internal calibration. Panel a shows the mass measurement errors across a wide *m/z* range from a spectrum internally calibrated using $[Na_{n+1}(COOH)_n]^+$ clusters with standard error (standard deviation of the mass measurement error) of 168 ppb. Spectra of very complex samples, such as combinatorial libraries, will likely contain some unresolved species. Peak overlaps negatively affect the mass measurement accuracy. Panel b shows the mass measurement errors for 524 resolved monoisotopic peptide peaks in the spectrum in Figure 1 after internal calibration. The standard error of the calibrant peptides was 273 ppb, and the largest errors were found in the middle of the *m/z* range, which is also the most crowded region of the spectrum. This differs from the normal behavior in FTICR, where mass measurement uncertainty increase with *m/z*.

Cleavage and side chain deprotection of the peptides was performed with a mixture of TFA/water/mercaptoethane 90/5/5 for 2.5 h. After the addition of 2% triethylsilane the peptide material was precipitated with 10 volumes of ether/pentane 50/50. The peptide material was dissolved in water/acetic acid 90/10 and lyophilized overnight.

Calculation of Theoretical Mass Spectra. The prediction of relative intensities within the isotopic envelope of a single chemical compound is straightforward as isotope effects on ionization, ion transfer, and detection are negligible for molecules larger than ~100 Da in most experimental situations. The deviations between measured and theoretical distributions are due to factors such as random sampling or counting errors (a finite, discrete number of ions are measured), chemical background, detector noise, and finally a difference between the actual isotope ratios in the sample and the values used in the calculations. The prediction of mass spectra for several compounds is more difficult, as chemical species vary considerably in their ionization efficiency. To complicate things further, this ionization efficiency also depends on extraneous variables such as the sample matrix, pH, and electrospray solvent composition. The ion transfer and detection typically also varies between ions of very different *m/z*. However, for very complex mixtures such as combinatorial libraries, there are on average many compounds per elemental composition. If the design of the library does not infer a trivial correlation between elemental composition and ionization efficiency, it may be possible to predict a direct analysis mass spectrum reasonably well by assuming equal ionization efficiency for all compounds in the library, as the differences in ionization between species will be averaged over many species at each measured *m/z*.

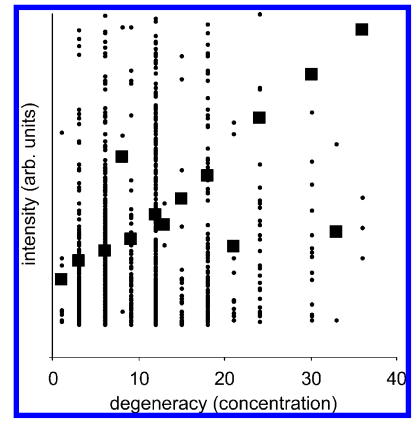

Figure 4. Average signal intensity (large squares) at each level of degeneracy calculated from 1535 observed monoisotopic and ^{13}C isotopic masses (points). In the 8,000-peptide library, the degeneracy varies from one for peptides with unique elemental compositions, such as GGGGLYNLVK, CCCGLYLVK, or WWWGLYLVK, to 36 (neutral elemental composition $C_{51}H_{85}N_{13}O_{15}$). The outliers are from rare numbers of degeneracy, averaging a small number of peak intensities. The degeneracy number corresponds to concentration in units of 250 nM. (The figure is cropped for clarity.)

Combinatorial library spectra were predicted with isotopic fine structure using the FFT-based approach[25] with oversampling,[26] adding each calculated peptide isotopic envelope to a vector of 2×10^6 data points, the index encoding the mass in mDa, corresponding to a bin size of 1 mDa. This is sufficient for accurately predicting broadband mode spectra with resolving powers up to $\sim 1 \times 10^6$ at *m/z* 1000. An existing MATLAB program (www.ms-utils.org/isotop_fs.html) was modified with an additional loop over the calculation of a single isotopic distribution. A more realistic spectrum can be simulated by convoluting the calculated discrete distribution with an experimental peak shape. However, the latter is typically dependent on *m/z* and instrument settings, and these must also be taken into account for simulating spectra covering a wide *m/z* range.

Mass Spectrometry and Data Analysis. The 20 synthesized combinatorial sub-libraries were resuspended in 50% MeOH, 50% H_2O, and 0.1% formic acid and pooled to a final concentration of 250 nM of each peptide, assuming uniform yield of the peptide synthesis. All mass spectra were acquired on a 15-T solariX™ FTICR mass spectrometer (Bruker Daltonics, Billerica, MA) in positive electrospray mode by direct infusion using the built-in syringe pump, a 100 μL syringe, and a flow rate of 120 μL/h. For each spectrum, 256 spectra of 2^{21} (~2 million) data points were

added, apodized (sinebell), and internally calibrated using calculated monoisotopic peptide masses and corresponding well-resolved peaks. Prior to the internal calibration, sodium formate was used for external calibration across a wide mass range from $[Na_{n+1}(COOH)_n]^+$ clusters. ApexControl version 3.0.0 build 72 (Bruker) was used for all acquisition and calibration. Peak lists and spectra were exported from DataAnalysis 4.0 build 234 (Bruker) for further analysis and comparison with predicted spectra using MATLAB 7.4.0 (The MathWorks, Natick, MA).

References and Notes

(1) Ellington, A. D.; Szostak, J. W. *Nature* **1990**, *346*, 818–822.
(2) Houghten, R. A. *Proc. Natl. Acad. Sci. U.S.A.* **1985**, *82*, 5131–5135.
(3) Furka, A.; Sebestyen, F.; Asgedom, M.; Dibo, G. *Int. J. Pept. Protein Res.* **1991**, *37*, 487–493.
(4) Houghten, R. A.; Pinilla, C.; Blondelle, S. E.; Appel, J. R.; Dooley, C. T.; Cuervo, J. H. *Nature* **1991**, *354*, 84–86.
(5) Lam, K. S.; Salmon, S. E.; Hersh, E. M.; Hruby, V. J.; Kazmierski, W. M.; Knapp, R. J. *Nature* **1991**, *354*, 82–84.
(6) Needels, M. C.; Jones, D. G.; Tate, E. H.; Heinkel, G. L.; Kochersperger, L. M.; Dower, W. J.; Barrett, R. W.; Gallop, M. A. *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 10700–10704.
(7) Schreiber, S. L. *Science* **2000**, *287*, 1964–1969.
(8) Hiemstra, H. S.; Duinkerken, G.; Benckhuijsen, W. E.; Amons, R.; de Vries, R. R.; Roep, B. O.; Drijfhout, J. W. *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94*, 10313–10318.
(9) Houghten, R. A. *Gene* **1993**, *137*, 7–11.
(10) Gonzalez-Vera, J. A.; Lukovic, E.; Imperiali, B. *Bioorg. Med. Chem. Lett.* **2009**, *19*, 1258–1260.
(11) El Oualid, F.; van den Elst, H.; Leroy, I. M.; Pieterman, E.; Cohen, L. H.; Burm, B. E.; Overkleeft, H. S.; van der Marel, G. A.; Overhand, M. *J. Comb. Chem.* **2005**, *7*, 703–713.
(12) Baine, M.; Georgie, D. S.; Shiferraw, E. Z.; Nguyen, T. P.; Nogaj, L. A.; Moffet, D. A. *J. Pept. Sci.* **2009**, *15*, 499–503.
(13) Guerrier, L.; Claverol, S.; Fortis, F.; Rinalducci, S.; Timperio, A. M.; Antonioli, P.; Jandrot-Perrus, M.; Boschetti, E.; Righetti, P. G. *J. Proteome Res.* **2007**, *6*, 4290–4303.
(14) Franz, A. H.; Liu, R.; Song, A.; Lam, K. S.; Lebrilla, C. B. *J. Comb. Chem.* **2003**, *5*, 125–137.
(15) Youngquist, R. S.; Fuentes, G. R.; Lacey, M. P.; Keough, T. *Rapid Commun. Mass Spectrom.* **1994**, *8*, 77–81.
(16) Thakkar, A.; Cohen, A. S.; Connolly, M. D.; Zuckermann, R. N.; Pei, D. *J. Comb. Chem.* **2009**, *11*, 294–302.
(17) Srebalus, B.; Hilderbrand, A. E.; Valentine, S. J.; Clemmer, D. E. *Anal. Chem.* **2002**, *74*, 26–36.
(18) Ludlow, R. F.; Otto, S. *J. Am. Chem. Soc.* **2008**, *130*, 12218–12219.
(19) Makarov, A.; Denisov, E.; Kholomeev, A.; Balschun, W.; Lange, O.; Strupat, K.; Horning, S. *Anal. Chem.* **2006**, *78*, 2113–2120.
(20) Marshall, A. G.; Hendrickson, C. L.; Jackson, G. S. *Mass Spectrom. Rev.* **1998**, *17*, 1–35.
(21) Demirev, P. A.; Zubarev, R. A. *Anal. Chem.* **1997**, *69*, 2893–2900.
(22) Nawrocki, J. P.; Wigger, M.; Watson, C. H.; Hayes, T. W.; Senko, M. W.; Benner, S. A.; Eyler, J. R. *Rapid Commun. Mass Spectrom.* **1996**, *10*, 1860–1864.
(23) Marshall, A. G.; Rodgers, R. P. *Acc. Chem. Res.* **2004**, *37*, 53–59.
(24) Marshall, A. G.; Rodgers, R. P. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 18090–18095.
(25) Rockwood, A. L.; Van Orden, S. L. *Anal. Chem.* **1996**, *68*, 2027–2030.
(26) Palmblad, M.; Buijs, J.; Håkansson, P. *J. Am. Soc. Mass Spectrom.* **2001**, *12*, 1153–1162.

CC9001235