BioSystems 109 (2012) 179–185

Contents lists available at SciVerse ScienceDirect

BioSystems

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

Periodic patterns in distributions of peptide masses

Shane L. Hubler[a,b,*], Gheorghe Craciun[a,c,*]

[a] Department of Mathematics, University of Wisconsin, Madison, USA
[b] Department of Chemistry, University of Wisconsin, Madison, USA
[c] Department of Biomolecular Chemistry, University of Wisconsin, Madison, USA

ARTICLE INFO

Article history:
Received 17 November 2011
Received in revised form 12 April 2012
Accepted 23 April 2012

Keywords:
Peptide databases
Mass distribution
Mass spectrometry
Periodic pattern

ABSTRACT

We are investigating the distribution of the number of peptides for given masses, and especially the observation that peptide density reaches a local maximum approximately every 14 Da. This wave pattern exists across species (e.g. human or yeast) and enzyme digestion techniques.

To analyze this phenomenon we have developed a mathematical method for computing the mass distributions of peptides, and we present both theoretical and empirical evidence that this 14-Da periodicity does not arise from species selection of peptides but from the number- theoretic properties of the masses of amino acid residues. We also describe other, more subtle periodic patterns in the distribution of peptide masses. We also show that these periodic patterns are robust under a variety of conditions, including the addition of amino acid modifications and selection of mass accuracy scale.

The method used here is also applicable to any family of sequential molecules, such as linear hydrocarbons, RNA, single- and double-stranded DNA.

1. Introduction

One of the most important methods for characterizing and sequencing proteins is based on measuring their masses using mass spectrometry. This field has enjoyed considerable growth in recent years due to the improved quality of the instruments (e.g. Scigelova and Makarov, 2006) and algorithms (Coon, 2009). In particular, peptides and proteins can be fragmented into smaller pieces that, when measured in a mass spectrometer, provide information about the structure of the original object.

One of the reasons that mass spectrometry is so successful for analyzing peptides is that peptides are linear; that is, they consist of a sequence of amino acids in a particular order. Under ideal circumstances, the fragmentation of a peptide gives rise to a series of singly charged ions whose masses are distributed in two series, one in the forward direction and one in the reverse, such that adjacent masses differ by the mass of a single amino acid. A plot of the relative abundances of these ions in a fragmented sample is known as a mass spectrum (Fig. 1). Good quality mass spectra have sharp peaks corresponding to masses present in the sample. Unfortunately, mass spectra usually are missing some of these peaks and have extra peaks.

Each line in the spectrum represents the mass-to-charge ratio of a peptide fragment. If we assume that each fragment is singly charged, then each line in the spectrum represents the mass of a fragment. One of the main goals of mass spectrometry in proteomics is to infer the sequence of peptides based on a mass spectrum.

This addition and removal of mass peaks means that there are often many possible peptides to match the spectrum. As a result there are many different scoring algorithms, scoring a peptide/spectrum pairing (Deutsch et al., 2008; Craig and Beavis, 2004; Eng et al., 1994; Geer et al., 2004; Perkins et al., 1999). One such scoring strategy is to compute a significance score based on the number of possible pairings with a certain score or better (Alves et al., 2007; Kim et al., 2009). These techniques require knowing the number of peptides within a specified mass accuracy of a given mass. However, the current technology requires the enumeration of all possible sequences in order to count them.

To identify alternative methods for counting peptide sequences, we (and others (Tipton et al., 2008; Mao et al., 2011; Yu et al., 2011)) investigated the distribution of peptides for given masses. In this investigation we noted, as the others have, a *periodic pattern* of approximately 14 Da. Note that our work here and in (Hubler and Craciun, 2012; Hubler, 2010) refers to ordered sequences, whereas (Tipton et al., 2008; Mao et al., 2011; Yu et al., 2011) refer to unordered sequences. For a general mathematical discussion of unordered sequences, see (Hubler and Craciun, submitted for publication). In this paper we also describe how this periodic pattern arises mathematically from the sequential addition of amino

* Corresponding authors at: Department of Mathematics, University of Wisconsin, Madison, USA.
E-mail addresses: slhubler@chem.wisc.edu (S.L. Hubler), craciun@math.wisc.edu (G. Craciun).

0303-2647/$ – see front matter © 2012 Elsevier Ireland Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.biosystems.2012.04.008

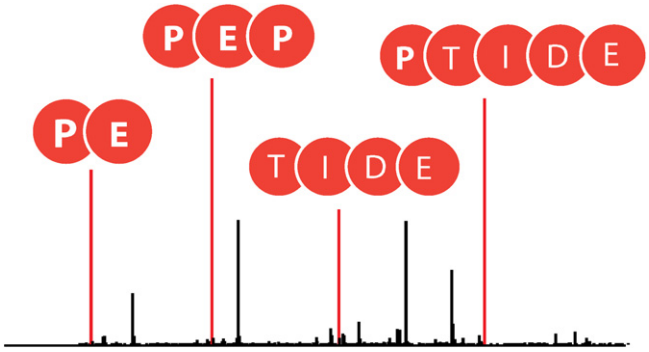

Fig. 1. Sample mass spectrum of a peptide fragment.

acids and investigate the robustness of this pattern and other, more subtle patterns. In addition, we point out that our mathematical analysis is applicable to any other family of sequential molecules, such as DNA, RNA, and linear hydrocarbons.

Finally, we note that the 14-Da periodic pattern, perhaps combined with sub-Dalton patterns, can be used to improve probability models of peptide distribution, as demonstrated in (Mao et al., 2011).

2. Wave patterns in databases

We are investigating the distribution of the number of peptides for given masses. Similar to (Tipton et al., 2008; Mao et al., 2011; Yu et al., 2011) we started our investigation with a histogram of the in silico digests of natural proteomes (yeast, human, and mouse), although we counted the number of ordered sequences instead of the number of unordered sequences (i.e. compositions). This led to the observation that peptide density reaches a local maximum approximately every 14 Da (Fig. 2). This wave pattern exists across species and enzyme digestion techniques (see Ram et al., 2005) for a detailed description of modern mass spectrometry techniques).

This pattern and periodicity exists regardless of species (e.g. yeast, human, or mouse), digest (e.g. Lys-C or Trypsin), and number of missed cleavages (MC), even though the absolute numbers change. This data was generated from in silico digests.

The intriguing 14-Da periodic pattern observed in Fig. 2 leads one to ask: did evolution select for proteins whose peptides exhibit this pattern? The counts of peptides from natural sources (yeast, human, and mouse) and those without selection (all possible peptides, with and without a large number of possible post-translational modifications) agree both in periodicity and phase (Fig. 3). Not only does this pattern exist when we ignore biological selection, but it also remains robust when we ignore the fact that most of the peptides in our biological samples include a lysine (due to digest) and even if we add 41 other masses to the mix (see the Supporting Table S2 for the list of 41 post-translational

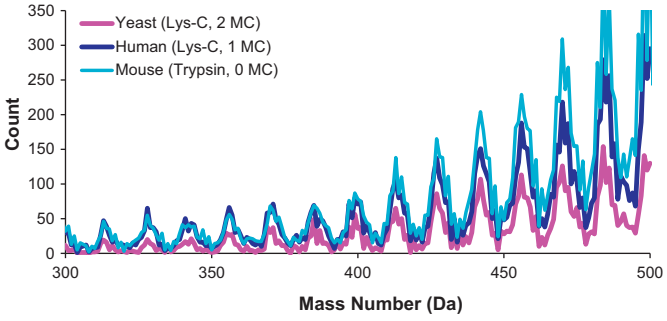

Fig. 2. Wave pattern found in peptide databases.

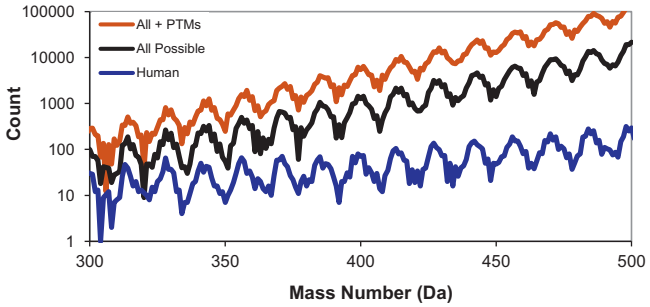

Fig. 3. Wave pattern is not caused by natural selection or by specific selection of amino acids.

modifications). In other words, this pattern must somehow arise from the masses of the amino acids themselves – the only common factor between the four scenarios.

The 14-Da wave pattern exists, both in relative magnitude and phase, when we include all theoretically possible peptides. The values represent number of peptides found per mass number (nominal mass). All + PTMs is a histogram including all possible peptides plus the inclusion of all 41 PTMs in Table S2. Note that the y-axis uses a log-scale, allowing us to compare the different accuracy levels of these objects.

3. Mathematical model

As a simple example, let us consider the special case where there are just two "amino acids" A and B with masses 1 and 2, respectively. In this case, the number of "peptides" of mass M can be partitioned into two groups: those ending in A and those ending in B. This leads to the following recurrence relation:

$$C(M) = C(M-1) + C(M-2) \tag{1}$$

which is the well-known recursion relation describing Fibonacci numbers (Fomin, 1988; Stanley, 1988). The initial conditions ($C(-1) = 0$, $C(0) = 1$) are different from the standard formulation of Fibonacci numbers, leading to a shift of the index. Table 1 shows the different types of solutions we use. Each of these representations has a different purpose, which we will discuss later.

We can generalize the recurrence relation of Eq. (1) for an arbitrary list of integer masses:

$$C(M) = \sum_{j=1}^{d} C(M - m_j) \tag{2}$$

where d is the number of amino acids (i.e. 20) and m_j is the nominal (integer) mass of the jth amino acid. Numerical sequences generated by recurrence relationships in the form of Eq. (2) are known as k-generalized Fibonacci numbers (Yang, 2008).

Table 1
Example of solutions to an ordered masses problem–masses of 1 and 2 (Fibonacci numbers).

Solution mode	Example solution
Sequence	1, 1, 2, 3, 5, 8,...
Recursion relation	$C(M) = C(M-1) + C(M-2)$
Exact closed-form solution	$C(M) = \left(\frac{5+\sqrt{5}}{10}\right)\left(\frac{1+\sqrt{5}}{2}\right)^M + \left(\frac{5-\sqrt{5}}{10}\right)\left(\frac{1-\sqrt{5}}{2}\right)^M$
Approximate formula	$C(M) = (0.723607)(1.618034)^M$
Sequence analysis	The dominant term increases exponentially (doubling time of 1.44042), while the other term has period 2 and decays exponentially (half-life of 1.44042).

Next we describe a theoretical approach to the periodicity question by determining an explicit mathematical formula for the number of peptides, $C(M)$, with a specific positive integer mass M.

3.1. General solution to sequence counting problem

Continuing our example of the Fibonacci numbers, note that the exact, closed-form solution to the well-known recurrence relation includes two terms summed together: the first and largest in magnitude is an exponential term while the second, being a negative number raised to an integer power, oscillates between positive and negative. In other words, the Fibonacci numbers increase exponentially with an oscillating term added to a pure exponential term.

The pattern found in this simple case turns out to be the pattern in general: all non-trivial sequence counting problems have a solution which is exponential in nature with a collection of periodic terms added, as described in the following theorem:

Theorem 1. *Let C, a function of integer mass M, count how many sequences of a finite set of integer masses have a combined mass* of *M. If the characteristic polynomial of Eq. (2) has no multiple roots then there exist real constants* $k, c_0, \ldots, c_k, r_0, \ldots, r_k, \theta_1, \ldots, \theta_k, \varphi_1, \ldots, \varphi_k$ such that

$$C(M) = c_0 r_0^M + \sum_{j=1}^{k} c_j r_j^M \cos(2\pi\theta_j M + \varphi_j) \tag{3}$$

and $r_0 \geq r_j > 0$ for $j = 1, \ldots, k$.

Eq. (3) follows from standard approaches to solving the difference equation, Eq. (2) (Balakrishnan, 1996; Cormen, 2009; Kelley and Peterson, 1991), and using $r-\theta$ notation and Euler's equation to convert the results into an exponential times a complex number. We note that, by Descartes' Rule of Signs, we can determine that there is exactly one positive root (which corresponds to the first, purely exponential term). Then, by grouping complex conjugates, we are left with exponentials times the sum on the right. Proving that the exponential term dominates the rest of the terms (i.e. that $r_0 \geq r_j > 0$) requires more mathematical machinery (Bergelson, 2003; Hardy and Littlewood, 1914) but generally follows from showing that the sum is negative infinitely often and, if the left term does not dominate the right term, the sum of the two must be negative infinitely often, which is a contradiction. See Hubler (2010) or Hubler and Craciun (submitted for publication) for a rigorous mathematical proof and an extension of this theorem.

Application of Theorem 1 is described below. It should be noted that all of the chemically motivated cases described in this paper have solutions of this form (i.e. their characteristic polynomials have no multiple roots).

3.2. Specific solution for peptide masses

By applying Theorem 1 and solving for the constants $c_0, \ldots, c_k$, $r_0, \ldots, r_k$, $\theta_1, \ldots, \theta_k$ and $\varphi_1, \ldots, \varphi_k$ we obtain the results shown in Fig. 4. The largest term, corresponding to the only positive root of the characteristic polynomial, has the $r-\theta$ coordinates in the complex plane of $(1.0285, 0^\circ)$, which correspond to a doubling time of approximately 24.7 Da.

The next largest terms, represented by the two red dots in Fig. 4(a), have $r-\theta$ coordinates in the complex plane of $(1.0255, \pm 25.2157^\circ)$, which correspond to a period of 14.2768 Da $(360^\circ/25.2157^\circ)$. The largest terms after *that*, $(1.0214, \pm 177.9690^\circ)$ have a period of 2.0228 Da. In other words, the third period shows an even/odd relationship that reflects the fact that many of the amino acids have an odd mass number.

The two blue dots correspond to the real roots, one of which is the dominating term and the other a negative root (with zero constant). The red dots correspond to the second largest terms, which account for most of the periodicity observed in Figs. 2 and 3.

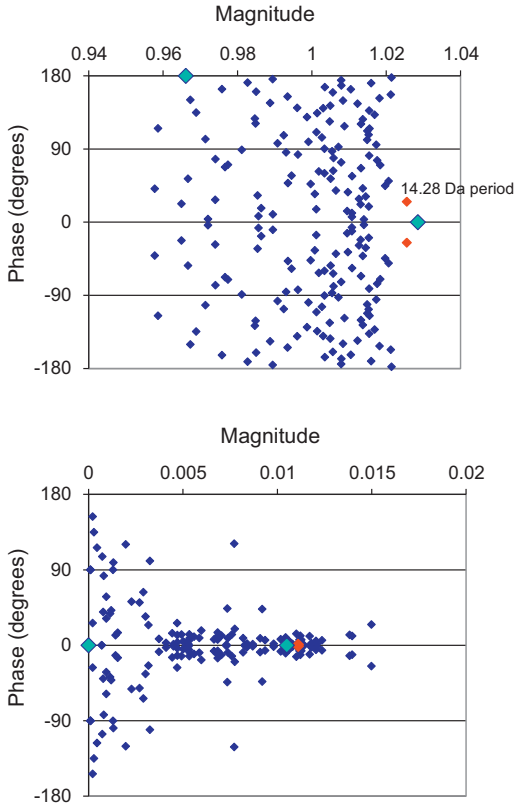

Fig. 4. Roots (top) and constants (bottom) of the characteristic polynomial for integer masses of amino acids. (For interpretation of references to colour in the sentence, the reader is referred to the web version of this article.)

Therefore, we can explain the 14-Da periodicity using this simple mathematical model.

Another consequence of this mathematical model is that it shows that there is more than one periodicity. The largest magnitude terms in Eq. (3) dominate the overall solution but many terms can play a role in its finer structure. Fig. 5 shows the relative contribution of each term. Of particular interest is the fact that the contribution of the first three terms dominates the solution beyond 750 Da. Fig. 6 shows how many terms are needed for adequate representation at various masses.

The relative contribution of each term is computed by looking at the magnitude of the term, ignoring the phase in the complex

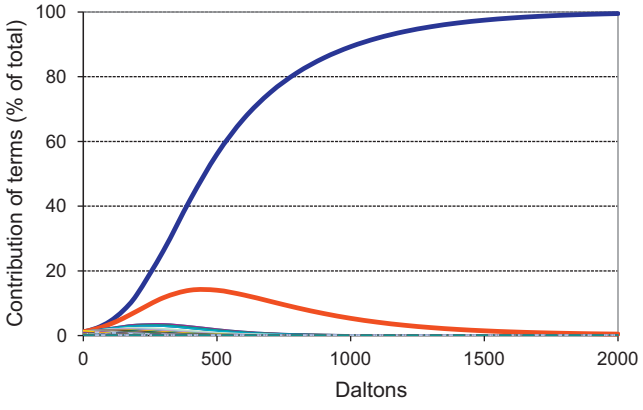

Fig. 5. Relative contribution of each term.

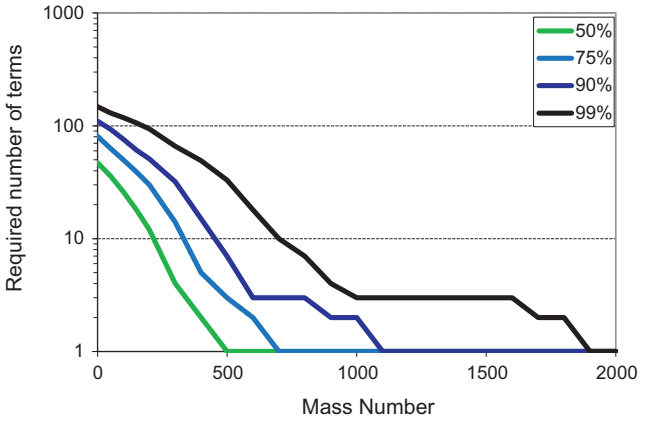

Fig. 6. Required number of terms to achieve proportion of the total number of sequences. (For interpretation of references to colour in the sentence, the reader is referred to the web version of this article.)

plane. The dominant blue curve shows the relative contribution of the term corresponding to the positive root of the characteristic polynomial. The relative magnitudes of all 186 terms are shown but most of them are insignificant, especially beyond 500 Da.

The percentage calculation is the same as measured in Fig. 5. Note the log scale in the *y*-axis.

Fig. 7 demonstrates this concept more explicitly. As the mass *M* gets larger, the approximation based on the first three terms improves (on a relative scale). The top panel of Fig. 7 also illustrates the second most common periodicity (2.02 Da) is strongly represented within the range of 150–200 Da.

Results are plotted in three different mass ranges: (top) 150–200 Da, (middle) 450–500 Da, and (bottom) 950–1000 Da. Note that, while the *x*-axis ranges are similar, the *y*-axis ranges are vastly different. Even though the *y*-axis ranges differ, we did not use a log scale, as we have on previous figures. This illustrates that the periodic patterns increase in absolute size (the wave increases as we look from left to right in each plot above). On the other hand, the relative size of the wave decreases (see Fig. 3).

We also considered other types of theoretical databases, including four different families of post-translational modifications (see Supporting Information).

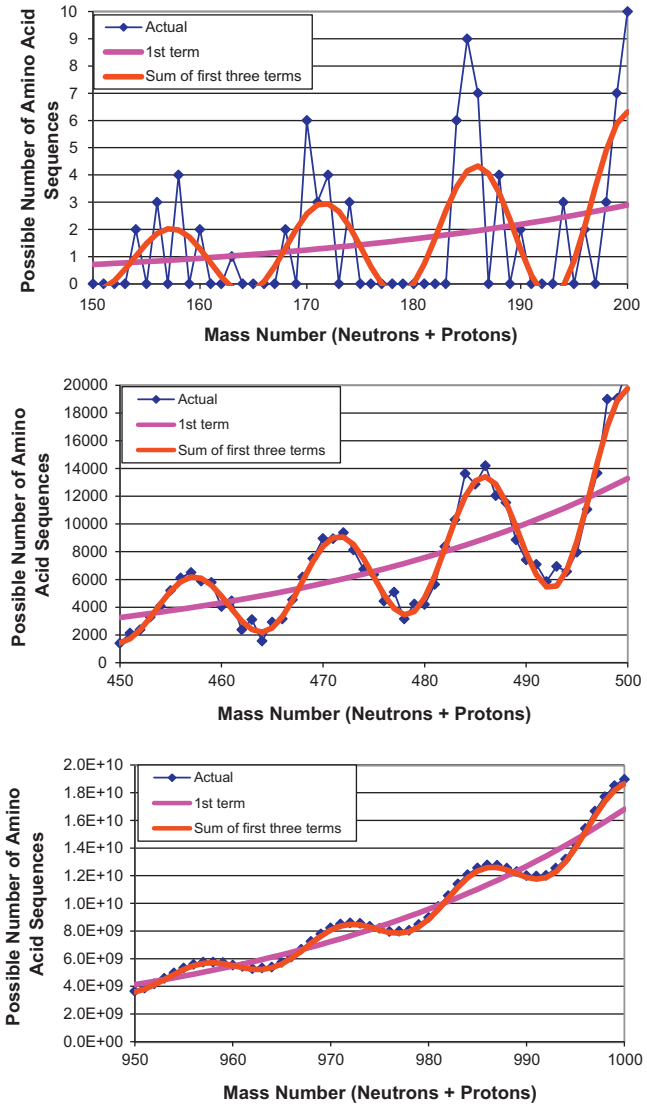

Fig. 7. Effectiveness of first three terms in approximating number of amino acid sequences.

4. Effect of mass accuracy on period

Other researchers have found that certain units of mass are special in the sense that the amino acids are close to integer values of those units and, indeed, 1 Da is remarkably close to such a unit of mass (1.000416 Da) (Alves and Yu, 2008). In principle the choice of units may play a role in forming the periodic pattern. After all, Fig. 3 was obtained by using mass numbers, the integer form of the true masses. While the nominal (integer) mass has physical meaning, specifically the number of neutrons and protons in the ion, mass spectrometry measures mass, not mass number. Could our choice of units generate the 14-Da pattern?

4.1. Accuracy effect on 14-Da pattern

Using the accurate mass values of the 20 amino acids, we created histograms of the number of sequences for bin sizes ranging from 0.1 Da to 3 Da; for bin size *b*, we divided a mass by *b* and rounded down to determine in which bin a mass fell. We also used a mass that was unsynchronized to the standard mass units: 2.718281 Da (an approximation of the irrational number *e*). We performed an analysis similar to that of the previous experiment.

Fig. 8 shows what happens when we round off using different accuracy levels. While the results for some scales appear more noisy (notably 0.1 Da), the 14-Da period appears in all

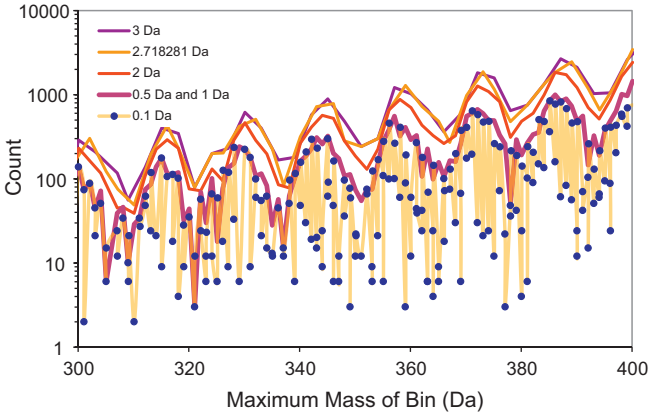

Fig. 8. Wave pattern exists regardless of bin size.

cases. So, indeed, scale affects the appearance of the histograms but it does not change the existence of the 14-Da period.

Each peptide mass was divided by the bin size (0.1–3) and rounded down. We used 2.718281 to illustrate the effect of selecting a scale unrelated to amino acids. The mass range was selected to be illustrative; this pattern is observable up to at least 1500 Da. Notice that the *y*-axis uses a log scale, allowing us to more easily see the periodic pattern.

4.2. Accuracy

The previous section describes a qualitative argument that the 14-Da period is robust under different accuracy levels. However, we would like to quantify that and also discuss the other periodicities implied by Theorem 1. One difficulty for quantifying periodicities across scales, though, is that the theorem assumes integer masses for the amino acids. In fact, this is a fundamental limitation of this technique; there are an infinite number of solutions to the analogue of the characteristic polynomial if we replace the masses with irrational numbers. However, it is possible to numerically solve the related polynomials where we multiply the masses by some common multiple. What happens to the periodicity as we increase the accuracy of our masses?

4.3. Periodicity analysis

Using Matlab 6.1 (MathWorks Inc., Natick, MA, 2000) we computed the roots of the characteristic polynomial and converted them to the form of Theorem 1. Next we verified that that there are no multiple roots, allowing us to apply Theorem 1. We also computed the periodicity for each term $Period(j) = 360^\circ/\theta_j$, and the respective λ_j's to verify that their values were comparable when the term's magnitude was greater than 1.

We performed the above periodicity analysis for accuracies ranging from 0.05 to 1 Da. For a given accuracy a, we divided each amino acid mass by a and rounded to the nearest integer. This led to characteristic polynomials of much higher degree. For example, an accuracy of 0.1 Da led to a polynomial of degree 1861 since the mass of the heaviest amino acid, tryptophan, has a residue mass of 186.08.

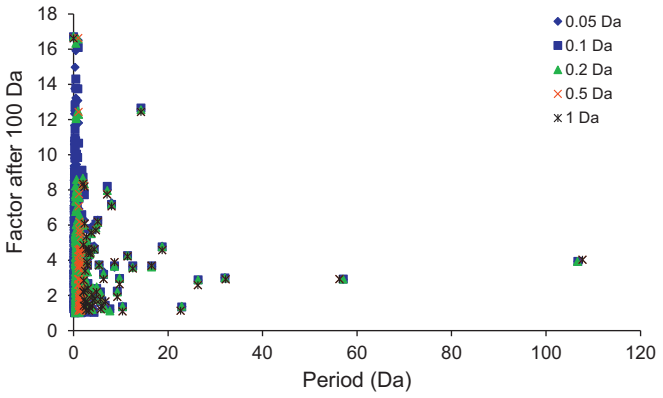

Fig. 9. Consistency of Periodicities by Scale.

This method of analyzing the effects of accuracy requires a change of variable from which we had to convert after finding the roots of the characteristic polynomial. In order to compare results across different accuracies we raised the magnitudes r_j to the power of $1/a$ and multiplied the calculated periodicity by a, since periodicity is inversely related to frequency.

To determine the relative importance of each term, we computed the factor $c_j r_j^M$ for masses ranging from 0 to 2000 Da. This represents the amplitude of the periodic terms and the absolute size of the non-periodic term in Eq. (3).

Fig. 9 shows that the higher periodicities are remarkably consistent across different accuracy levels, both in period and magnitude; we find that the 14-Da period is 14.2795 ± 0.0034 Da if we treat the four scales as four separate random trials. Also, what appears to be noise in Fig. 8 contains useful information – we now know that there are a variety of periodicities that exist regardless of scale (assuming that it is less than 1), most notably 106.6 Da, 56.2 Da, 32.2 Da, 18.8 Da, and 14.28 Da. Only the last is easily observable in the original data. However, the periodic patterns present at finer scales,

Table 2
Factors and remainders for amino acids.

Amino Acid	Mass	Zero of char. poly.		CH_2		CH_2 (Nominal)	
		Divisor					
		14.28		14.02		14	
		Factor	Remainder	Factor	Remainder	Factor	Remainder
Glycine	57	4	*0.12	4	*−0.94	4	*−1.00
Alanine	71	5	*0.40	5	*−0.92	5	*−1.00
Serine	87	6	*−1.32	6	−2.91	6	−3.00
Proline	97	7	2.96	7	*1.11	7	*1.00
Valine	99	7	*0.96	7	*−0.89	7	*−1.00
Threonine	101	7	*−1.04	7	−2.89	7	−3.00
Cysteine	103	7	−3.04	7	−4.89	7	−5.00
Iso-leucine	113	8	*1.24	8	*−0.87	8	*−1.00
Leucine	113	8	*1.24	8	*−0.87	8	*−1.00
Asparagine	114	8	*0.24	8	−1.87	8	−2.00
Aspartic Acid	115	8	*−0.76	8	−2.87	8	−3.00
Glutamine	128	9	*0.52	9	−1.86	9	−2.00
Lysine	128	9	*0.52	9	−1.86	9	−2.00
Glutamic Acid	129	9	*−0.48	9	−2.86	9	−3.00
Methionine	131	9	−2.48	9	−4.86	9	−5.00
Histidine	137	10	5.80	10	3.16	10	3.00
Phenylalanine	147	10	−4.20	10	−6.84	11	7.00
Arginine	156	11	*1.08	11	−1.83	11	−2.00
Tyrosine	163	11	−5.92	12	5.19	12	5.00
Tryptophan	186	13	*−0.36	13	−3.80	13	−4.00

such as periods around 2, 1, and 0.5 Da are common and strong for higher accuracy data.

Each amino acid mass was divided by the scale and rounded to the nearest integer. The periods θ_j and factors $c_j r_j^M$ were then calculated as described in the text. We used the size of the factor at 100 Da, i.e. $c_j r_j^{100}$, to compare all of the different accuracy levels and to visualize the relative amplitudes. We left out all the points corresponding to magnitudes r_j less than one, since they do not influence the final count significantly.

5. Why 14.28 Da?

While the solutions to the characteristic polynomial are concrete numbers with which we can describe our results, it is not very satisfying to explain that the 14.28-Da periodicity arises because it is the zero of a polynomial. Is there some chemical significance to 14.28? It is tempting to simply pay attention to the integer part and recognize that the nominal mass of CH_2 is 14. After all, CH_2 is a chemical component that is seen quite often in amino acid structure and, indeed, in organic molecules in general. In fact, many of the amino acids differ by exactly the mass of CH_2.

However, the fractional part, 0.28, is troubling. Table 2 shows that this is not a matter of round-off error; multiples of 14.28 are remarkably close to several of the masses, as compared to other candidates (14 and 14.02). In fact 70% of the amino acids fall within 10% of a multiple of 14.28, whereas only 30% of the amino acids are close to multiples of either 14.01 or 14. Note that we did not take the phase into account here; the phase on the 14.28-Da period is small, allowing us to ignore it for this analysis.

We should also point out that a period around 14 Da is reasonable since the amino acid residues can all be constructed (with rearrangement) from five groupings of elements, a kind of elemental basis: C, CH_2, NH, O, and S, with nominal masses of 12, 14, 15, 16, and 32, respectively. Since S is only in two of the twenty amino acids, it is reasonable to suppose that the other four constituent components would have an effect in a periodic pattern arising from objects made up of this basis.

6. Conclusions

We have explained the observed 14-Da periodicity in the histogram of masses of natural peptides. We proved that a better approximation of this period is 14.28 Da and that it arises from mathematical properties of the amino acids masses.

This phenomenon occurs even if we vary species, cleavage enzyme (not shown), scale, or include of post-translational modifications. It also exists if we include every possible amino acid sequence, rather than just the ones found in nature. In other words, the periodic pattern is not a by-product of natural selection.

We can calculate an arbitrary number of periodicities but the larger periodicities remain fairly robust regardless of the scale of our approximation. While a periodicity of exactly 14 Da would be very satisfying from a chemical point of view (when working with integer masses), the actual periodicity is closer to 14.28 Da. We suggest an explanation of this result by noting that many more amino acid masses are closer to being integer multiples of 14.28 than to 14.01 (the mass of CH_2) or 14 (the nominal mass of CH_2). However, the fact that this number is probably irrational (i.e. not a ratio of two integers) means that local maxima in the histograms will shift in a quasi-periodic pattern. This is consistent with the observation by Yu et al. (2011) that the distance between adjacent peaks varied from 14 to 16 Da.

In addition to information about the periodic pattern, our analysis also provides useful information about the size of a database: we only need the first term to estimate how large a database should be in order to list all sequences up to a particular mass. Another way to look at this is to consider the largest mass M_{max} that will fit into a database of a particular size. Note that M_{max} is similar for different conditions, such as a large number of modifications (see Supporting Information, Table S-3). The exponential nature of these problems tells us that, no matter how much storage space we have, a database storing all potential sequences will not contain masses much beyond 600 Da. It also tells us that most of the sequences will have masses within 26 Da of the cut-off.

The methods described in this paper can be used at various scales and allow for considering masses at sub-Dalton resolution as long as all masses under consideration are integer multiples of the same common unit.

Knowing how many peptide sequences exist at particular masses may be useful for estimating the probability that a particular novel sequence comes from a natural source (Samuelsson et al., 2004). Furthermore, all of these techniques are applicable to other settings such as the addition of more masses (e.g. post-translational modifications), RNA, or DNA fragment masses, or even linear hydrocarbons.

Acknowledgements

We thank Joshua Coon and Alicia Williams for very useful discussions and comments. We would also like to thank the anonymous reviewers for very helpful suggestions. SLH was supported by NLM training grant Computation and Informatics in Biology and Medicine 5T15LM007359. GC and SLH were supported by NSF DBI-0701846.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.biosystems.2012.04.008.

References

Alves, G., Yu, Y-K., 2008. Statistical characterization of a 1D random potential problem—with applications in score statistics of MS-based peptide sequencing. Phys. A: Stat. Mech. Appl. 387, 6538–6544.
Alves, G., Ogurtsov, A.Y., Wu, W.W., Wang, G., Shen, R.F., et al., 2007. Calibrating e-values for MS2 database search methods. Biol. Direct 2.
Balakrishnan, V.K., 1996. Introductory Discrete Mathematics. Dover Publications, New York, xiv, 236 p.
Bergelson, V., 2003. Minimal idempotents and ergodic Ramsey theory. In: Bezuglyi, S., Kolyada, S. (Eds.), Topics in Dynamics and Ergodic Theory. Cambridge University Press, pp. 8–39.
Coon, J.J., 2009. Collisions or electrons? Protein sequence analysis in the 21st century. Anal. Chem. 81, 3208–3215.
Cormen, T.H., 2009. Introduction to Algorithm. The MIT Press, Cambridge, MA.
Craig, R., Beavis, R.C., 2004. TANDEM: matching proteins with tandem mass spectra. Bioinformatics 20, 1466–1467.
Deutsch, E.W., Lam, H., Aebersold, R., 2008. Data analysis and bioinformatics tools for tandem mass spectrometry in proteomics. Physiol. Genomics 33, 18–25.
Eng, J.K., Mccormack, A.L., Yates, J.R., 1994. An approach to correlate tandem mass-spectral data of peptides with amino-acid-sequences in a protein database. J. Am. Soc. Mass Spectrom. 5, 976–989.
Fomin, S.V., 1988. Generalized Robinson–Schensted–Knuth correspondence. J. Math. Sci. 41, 979–991.
Geer, L.Y., Markey, S.P., Kowalak, J.A., Wagner, L., Xu, M., et al., 2004. Open mass spectrometry search algorithm. J. Proteome Res. 3, 958–964.
Hardy, G., Littlewood, J., 1914. Some problems of diophantine approximation (Part I). Acta Math. 37, 155–191.
Hubler, S.L., 2010. Mathematical Analysis of Mass Spectrometry Data, Ph.D. Thesis. Madison, WI: University of Wisconsin-Madison.
Hubler, S., Craciun, G., 2012. Mass distributions of linear chain polymers. J. Math. Chem., doi:101007/s10910-10012-19983-z.
Hubler, S.L., Craciun, G. Counting chemical compositions using Ehrhart quasi-polynomials. J. Math. Chem., in press.
Kelley, W.G., Peterson, A.C., 1991. Difference Equations: An Introduction with Applications. Academic Press, Boston, xi, 455 p.

Kim, S., Bandeira, N., Pevzner, P.A., 2009. Spectral profiles, a novel representation of tandem mass spectra and their applications for de novo peptide sequencing and identification. Mol. Cell. Proteomics 8, 1391–1400.

Mao, Y., Tipton, J.D., Blakney, G.T., Hendrickson, C.L., Marshall, A.G., 2011. Valence parity to distinguish c' and $z_\bullet$ ions from electron capture dissociation/electron transfer dissociation of peptides: effects of isomers, isobars, and proteolysis specificity. Anal. Chem. 83, 8024–8028.

Perkins, D.N., Pappin, D.J.C., Creasy, D.M., Cottrell, J.S., 1999. Probability-based protein identification by searching sequence databases using mass spectrometry data. Electrophoresis 20, 3551–3567.

Ram, R.J., Verberkmoes, N.C., Thelen, M.P., Tyson, G.W., Baker, B.J., et al., 2005. Community proteomics of a natural microbial biofilm. Science 308, 1915–1920.

Samuelsson, J., Dalevi, D., Levander, F., Rognvaldsson, T., 2004. Modular, scriptable and automated analysis tools for high-throughput peptide mass fingerprinting. Bioinformatics 20, 3628–3635.

Scigelova, M., Makarov, A., 2006. Orbitrap mass analyzer—overview and applications in proteomics. Proteomics 6, 16–21.

Stanley, R.P., 1988. Differential posets. J. Am. Math. Soc. 1, 919–961.

Tipton, J.D.M., Hendrickson, Y.C.L., Marshall, A.G., 2008 December 7–10. Utility of the Valence Parity Rule for ECD/AIECD FT-ICR MS: H-dot Atom Transfer, Isobars, and Isomers for Peptide Analysis , Madison, WI.

Yang, S.L., 2008. On the *k*-generalized Fibonacci numbers and high-order linear recurrence relations. Appl. Math. Comput. 196, 850–857.

Yu, L., Xiong, Y-M., Polfer, N.C., 2011. Periodicity of monoisotopic mass isomers and isobars in proteomics. Anal. Chem. 83, 8019–8023.