



APPLICATION NOTE

***pepgrep*: A Tool for Peptide MS/MS Pattern Matching**

Igor Chernukhin *

School of Biological Sciences, University of Essex, Wivenhoe Park, Colchester CO4 3SQ, UK

Received 18 October 2012; revised 21 December 2012; accepted 19 February 2013

Available online 16 March 2013

KEYWORDS

Collision-induced dissociation;
 Tandem MS;
 Protein database search;
 Peptide identification

Abstract Typically, detection of protein sequences in collision-induced dissociation (CID) tandem MS (MS2) dataset is performed by mapping identified peptide ions back to protein sequence by using the protein database search (PDS) engine. Finding a particular peptide sequence of interest in CID MS2 records very often requires manual evaluation of the spectrum, regardless of whether the peptide-associated MS2 scan is identified by PDS algorithm or not. We have developed a compact cross-platform database-free command-line utility, *pepgrep*, which helps to find an MS2 fingerprint for a selected peptide sequence by pattern-matching of modelled MS2 data using Peptide-to-MS2 scoring algorithm. *pepgrep* can incorporate dozens of mass offsets corresponding to a variety of post-translational modifications (PTMs) into the algorithm. Decoy peptide sequences are used with the tested peptide sequence to reduce false-positive results. The engine is capable of screening an MS2 data file at a high rate when using a cluster computing environment. The matched MS2 spectrum can be displayed by using built-in graphical application programming interface (API) or optionally recorded to file. Using this algorithm, we were able to find extra peptide sequences in studied CID spectra that were missed by PDS identification. Also we found *pepgrep* especially useful for examining a CID of small fractions of peptides resulting from, for example, affinity purification techniques. The peptide sequences in such samples are less likely to be positively identified by using routine protein-centric algorithm implemented in PDS. The software is freely available at <http://bsproteomics.essex.ac.uk:8080/data/download/pepgrep-1.4.tgz>.

Introduction

Collision-induced dissociation (CID) tandem MS (MS2) became the front-line method in peptide analysis that is widely

used in discovery (shotgun) proteomics [1–3] and retains its leading position among other MS-conducted peptide-sequencing methods [4]. This is mainly because of its stable display of fragmentation pattern of y- and b-ions that helps to generate a sequence tag, which, in turn, leads to confident identification of peptide sequences [5,6]. Mascot [7], SEQUEST [8] and TANDEM [9] became the popular software for protein identification based on MS2-to-Peptide and Peptide-to-Protein scoring algorithms. Protein database search (PDS) method relies entirely on a premise that the computed theoretical data matches to scanned ion mass spectrum and, in reality, this works only for a relatively small (5–25%) portion of the spectra acquired during a single LC–MS run. The biggest ‘silent’

* Corresponding author.

E-mail: igorc@essex.ac.uk (Chernukhin I).

Peer review under responsibility of Beijing Institute of Genomics, Chinese Academy of Sciences and Genetics Society of China.



Production and hosting by Elsevier

part of acquired MS2 data is believed to contain a significant part of the ‘peptide-looking’ spectra that is usually left undiscovered by PDS. The reason for this is thought to be associated with (i) products of post-translational modifications (PTMs), (ii) missed cleavages, (iii) non-specific cleavages which may occur during protein breakdown and (iv) proteins whose sequences are not present in the searched database. Although modern search engines allow users to supply dozens of PTM-related mass offsets for PDS as additional options for fixed and/or variable PTMs, practically this is limited to simple amino acid side additions that may appear, for example, during oxidation and products of *in vitro* derivatization reactions. Input with a larger number of PTMs would be infeasible for standard computing architectures commonly used in research labs, due to elevated computational complexity in the search algorithm. Also this will lead to an increase in the number of false-positive estimations. Nevertheless, the above limitations are deemed to be not crucial for routine PDS protein discovery, because the majority of proteins will always yield some amount of PTM-free compounds during protein digest, which will be sufficient to satisfy PDS-scoring algorithm for confident identification of a protein.

Finding a particular peptide sequence in CID MS2 could be important for some applications that use a targeted approach like commonly used selected reaction monitoring-MS (SRM-MS) [10], or its deviation, multiple reaction monitoring-MS (MRM-MS) [11], technology. These may also include the selected detection of certain peptide sequences after affinity capture using unique PTMs as ligands. In either case, the successful surveying for the searched sequences will rely entirely on the PDS protein-centric scoring algorithm, which can return ambiguous results regardless of whether it is used with protein sequences or just a few peptides as a database. This may require more meticulous analysis of the tandem spectra which is typically associated with manual spectrum interpretation. However the large volumes of MS2 datasets produced by modern tandem MS instruments make manual examination very inconvenient and thus an automated database-free peptide-centric algorithm for fast matching and viewing of a sequence pattern in MS2 scans is thought to be beneficial. We found that using the Peptide-to-MS2 modelling algorithm supplemented with the parallel complementary tests for decoy sequences does produce good results for surveying a specified peptide sequence pattern in CID MS2 spectra.

Here we describe a simple database-free approach, *pepgrep*, for searching peptide pattern matching in CID MS2 data files that were obtained from popular types of MS tandem instruments. The name *pepgrep* is assigned by analogy to a popular POSIX utility *grep*. The matched MS2 spectra can be displayed using graphical API or saved to a file for further visual examination, along with the matched MS2 scan data that can be copied to a separate file. Optionally, it can include a number of mass offsets for amino acid residues related to specific PTMs.

Results

The ability of Peptide-to-MS2 algorithm to score identically to PDS algorithm for the same MS2 pattern was estimated by comparing the results produced by both popular search engines Mascot and TANDEM with those by *pepgrep*. For this

purpose, we created a diagnostic MS2 dataset; a subset of the spectra was selected based on high confidence scoring by both TANDEM and Mascot search engines and expert-conducted visual quality estimation. MS2 data were generated by the Q-TOF instrument (“Premier”, Waters) with routine LC-MS run with a tryptic protein digestion. The representative subset was subsequently analysed with *pepgrep* and found that all PDS-selected MS2 ions were also among the high-scored peptide patterns matched by *pepgrep* (data not shown). This suggests that Peptide-to-MS2 algorithm is cross-compatible with PDS and can be implemented for peaking up and visualizing a single peptide pattern match among MS2 spectra during iteration of the dataset.

In another instance, we used a combination of TANDEM and *pepgrep* to analyse the products of co-immunoprecipitation (co-IP) reaction of mammalian cell proteins. The enriched co-IP protein band of interest was extracted from a Coomassie stained gel and subjected to in-gel tryptic digestion. The CID MS2 spectra were obtained by using an ion-trap instrument (Esquire, Bruker) during standard LC-MS run. MS2 identification with TANDEM revealed a number of proteins. Among them, one was of particular interest, which is ATP-dependent RNA helicase A (Swiss-Prot accession No. Q08211). Using *in silico* digestion, we generated all tryptic peptides for this protein with the number of missed cleavages equal to two and a set of expected PTMs allowed into the search. Analysis with *pepgrep* displayed the presence of MS2 spectra matching a significantly larger number of peptides than those identified during TANDEM search (Table 1, see detailed spectra in supplementary Figure S1). All of the peptide-associated MS2 patterns resembled a good profile for both y- and b- ions estimated by visual examination.

Finally, we evaluated *pepgrep* in finding expected sequence patterns in MS2 scans generated by ion trap LC-MS analysis of affinity-captured peptides. In this experiment, we affinity-purified a peptide fraction attached to poly-ADP-ribose (PAR) using a boronic acid affinity media. The N-terminus of CCCTC-binding factor (CTCF), a transcriptional factor [12], is known to carry the PAR moiety; however, there is no method available to directly detect the peptide sequences carrying this modification. The affinity purification procedure included binding, washing and *in situ* mild alkaline hydrolysis, which led to detachment of the peptides from the modification and their elution from the media. The eluted peptides were subjected to the CID MS2 scan and the output was analyzed with *pepgrep* matching *in silico* digested peptides from N-terminus of CTCF. The expected sequence ‘VVGNMKPPKPTKIK’ gave a good pattern match to one of the analyzed spectra with a confident score. No matching peptides were found when using either Mascot or TANDEM.

All datasets used for these examples are supplied with *pepgrep* package for testing purposes.

Discussion

pepgrep offers a quick and friendly way to examine MS2 output data for the presence of a target peptide by matching MS2 spectrum patterns of modelled peptide sequence. It is primarily designed for use with MS2 datasets generated by CID MS scan of a peptide mix subjected to targeted proteomics. This may include products of *in vitro* derivatization or affinity purification

Table 1 List of identified peptides from ATP-dependent RNA helicase A after *pepgrep* scrutiny of TANDEM - generated MS2 dataset

TANDEM		Pepgrep		
Peptide sequence	Score	Peptide sequence	Score	FDR value
AAEC(57.02)NIVVTQPR	47	AAEC(42.000)N(28.000)IVVTQPR	27.47	4.737652e-03
AAMEALVVE(48.13)VTK(348.08)	25.9	AAEC(57.000)NIVVTQPR	138.02	1.110223e-16
AGY(80.0)GAGVGGGYR(48.13)	27.3	AGYGAGVGGGY(80.000)R	27.15	4.737652e-03
DFVNYLVR	38.2	DAQSNAAR	31.39	3.512739e-05
DINTDFLLVVL	43.3	DFVNYLVR	94.78	1.110223e-16
ELDALDANDELTPRGR	54.3	DIN(15.900)TDFLLVVL	24.18	1.440836e-02
GPSGGYR(48.13)GSGGFQR(48.13)	29.5	DINTDFLLVVL	28.43	3.213288e-03
ISAVSVAER	27.1	DVVQAYPEVR	30.69	3.397741e-04
LAAQ(0.984)SC(57.02)ALSLVR	36.5	ELDALDANDELTPRGR	63.33	1.151720e-08
LAAQSC(57.02)ALSLVR	41.2	FCE(132.000)HKR	14.52	2.069473e-02
LAQFEPSQR	34.9	GAYGTGY(80.000)FGQGR	28.93	6.426576e-03
LGGIGQFLAK	41.4	ILAKLPIEPR	22.87	1.003080e-02
NFLYAWC(16.0)(57.022)GKR	25.8	KDAQ(14.000)SNAAR	20.63	1.205977e-02
Q(-17.026)PAIISQLDPVNER	48.4	KILTTEGR	23.15	1.663446e-03
Q(-17.02)LYHLGVVEAYSGLTK	49.2	KKEGETVEPYK	19.49	4.936857e-02
QPAIISQLDPVNER	30.9	LETHMTPM(16.000)FR	20.98	3.537777e-02
RGGGRGAYGTGYFG	33.6	LGGIGQFLAK	38.14	7.610913e-06
TTQVPQFILDDFIQNDR	50.4	LN(16.000)ATLR	46.51	2.902661e-09
VFDVPVGVTK	38.6	LNQ(14.000)YFQK	16.68	1.734483e-02
VQ(0.984)SD(48.13)GQ(0.984)IVLVD(48.13)D(48.13)WIK(348.08)	32.4	LPIEPRFGK	36.38	7.687045e-06
VQ(0.984)SDGQIVLVD(48.13)D(48.13)WIK(348.08)	36.5	MARYDN(80.000)GYRR	33.73	4.120412e-04
YPSPFVFGKEK	42.3	MGEEAE(132.000)IR	26.27	9.391127e-04
YTQVGPDPHNR	27.1	MTPS(146.000)YEIR	16.85	3.195733e-02
		NFLY(80.000)AWCGKR	19.62	3.071417e-02
		NFLYAWCGKR	25	4.520593e-03
		QLGRRIFAR	18.97	2.611949e-02
		RKMT(162.000)PSYEIR	33.21	8.507759e-05
		SFIAE(14.000)MTIYIK	20.96	3.537777e-02
		SFIAEMTIYIK	39.53	1.193067e-05
		TNLEQRK	19.72	4.677653e-03
		TPLHEIALSIK	35.76	7.820860e-05
		TTQVPQFILDDFIQNDR	47.62	3.866538e-05
		VAFERGEPEGK	74.54	9.292567e-14
		VFDVPVGVTK	46.31	3.756522e-07
		YDNGSGYR	17.68	2.164263e-02
		YPSPFVFGKEK	53.82	9.982491e-09
		YTQVGPDPHNR(0.900)R	26.84	1.966587e-03

Note: The protein has been identified by TANDEM with ‘expect’ value -79.8 using UniProtKB/Swiss-Prot database release 2012_03. The following mass modifiers for amino acid residues were supplied for TANDEM search: 15.994915@M, 15.994915@W, 0.9848@N, 0.9848@Q, 31.98983@M, 31.98983@W, 162.0528@N, 162.0528@T, 162.0528@K, 80@S, 80@T, 80@Y, 14.0157@C, 14.0157@K, 14.0157@R, 14.0157@H, 14.0157@D, 14.0157@E, 14.0157@N, 14.0157@Q, 28.0314@C, 28.0314@K, 28.0314@R, 28.0314@H, 28.0314@D, 28.0314@E, 28.0314@N, 28.0314@Q, 42.0471@C, 42.0471@K, 42.0471@R, 42.0471@H, 42.0471@D, 42.0471@E, 42.0471@N, 42.0471@Q, 57.0518@C, 71.0167@C, 42.0106@K, 42.0106@S, 0.984@Q, 0.984@C, 0.984@N, 18.010565@H, 18.010565@R, 15.9949@P, 15.9949@K, 15.9949@D, 15.9949@N, 15.9949@C, 348.08@K, 348.08@R, 132@R, 132@N, 132@E, 132@D, 78.13@R, 78.13@N, 78.13@E, 78.13@D, 48.13@R, 48.13@N, 48.13@E, 48.13@D, 146.057@S. *pepgrep* annotated spectra for the identified peptides are displayed in Figure S1.

using PTM-specific ligands, as the module can incorporate a large variety of amino acid mass-modifying values into the computation algorithm. The matched spectrum can be instantly visualized during the search time or saved as an image and data file for later examination. Tests with a vast range of samples of MS2 spectra from the PeptideAtlas repository (<http://www.peptideatlas.org>) revealed that the complementary scoring using decoy peptide sequences helps to validate the searched sequence among large number of noisy spectra. This approach also allows quick and confident detection of targeted sequences in chimeric spectra.

pepgrep can be easily plugged into targeted proteomic workflows as an additional module, which could help, for example, to select targeted sequences for quantitative SRM-MS or a similar experimental design where MS2 detection of

a particular peptide sequence is required. The difference in algorithms between PDS and *pepgrep* is illustrated in more details in **Figure 1**. While the current version of *pepgrep* is designed only for the analysis of CID datasets by using y/b-ion fragmentation patterns, it can be easily converted into c/z-ion matching module by modifying *pepgrep* code and thus making it possible to be used for other types of fragmentation like electron transfer dissociation (ETD)-MS.

Although *pepgrep* was primarily designed for use with targeted proteomics, its ability to use a greater number of mass modifiers in pattern matching can also be implemented in shotgun proteomics in a “post-discovery” stage. Identification of PTM-modified peptides remains a challenging task in discovery proteomics despite significant progress made in the development of peptide tandem mass spectrometry during the last

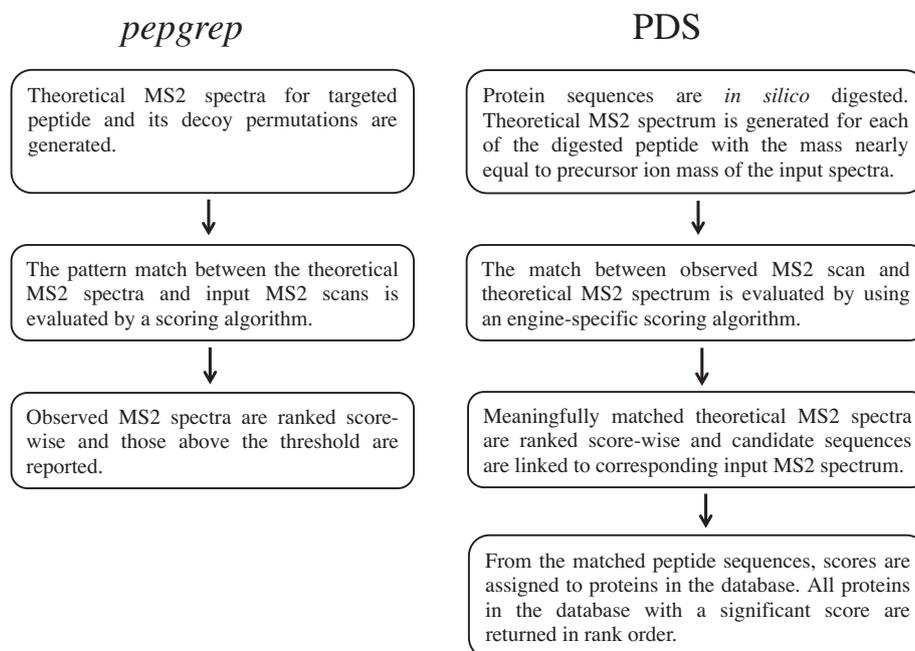


Figure 1 The flowchart outlines key steps in the search algorithms of *pepgrep* in comparison to a PDS engine

pepgrep evaluates a match of single selected peptide against all input MS2 spectra, whereas PDS algorithm scores a match of each MS2 spectrum against all peptide sequences in database.

decade and emerging new software products with new concepts for prediction and detection of PTMs. Numerous studies in shotgun proteomics showed that standard PDS alone cannot solve this task and should be reinforced with other methods and algorithms [13]. From the software point of view, there are several models that have been tested in proteomics workflows, which have been successful at detecting a broad range of PTMs. For example, PDS search can be successfully combined with *de novo* sequencing algorithm for filtering databases based on sequence tags as it was implemented in the *InsPecT* project [14]. Another example illustrates the use of TANDEM in ‘two-pass’ workflow [9], where the first discovery stage with routine PDS was followed by the second search of a mini-database generated from proteins identified in the discovery stage with a broad range of PTM-related mass offsets included. Similarly, *pepgrep* can be used for searching the patterns of potentially-modified peptides selected from a subset of peptides identified during a routine PDS run. It is crucial that the MS2 dataset contains good quality ion fragmentation spectra with low noise to meet criteria of complementary tests using decoy sequences.

The use of a cluster computing environment with *pepgrep* makes it easier to handle large MS2 data files that are usually generated by LC–MS run. The described approach provides an easy solution for comprehensive analysis of CID MS2 datasets.

Materials and methods

pepgrep was built in *Python* programming language. The module implements the *numpy* numeric library in all MS2 manipulation routines and also uses *numpy* array implementation for structured data handling and manipulation. There are four class modules that provide functionality to *pepgrep*, which

handle spectrum allocation, MS file input/output (IO) manipulations and MS2 spectrum routines, and build MS2 spectral models for peptide sequence with different types of fragmented ions.

Usage

```
pepgrep.py [-options or @OptionFile] <peptide sequence | -file file with peptide sequences> <input ms2 dataset> .
```

Options

Options can be supplied with a command line or with a file. Command line options are the following:

-ferr (digit)	mass tolerance for fragment ion
-perr (digit)	mass tolerance for peptide
-it (digit)	intensity threshold. The level of intensity above which peaks are considered significant. This option should accompany ‘thr’ option
-thr (Y N)	spectrum thresholding. This should be considered for quadrupole-time-of-flight (Q-TOF) or ion trap types of data
-filter (Y N)	turns on filtering the first three tryptic y-ions. This makes sense only for Q-TOF type of spectra
-pw (digit)	predefined peak width
-pmmixed (Y N)	turns on or off peptide mass validation. When turned on, <i>pepgrep</i> will select only spectra, for which both fragment and precursor masses of simulated and tested MS2 will match
-fdr (digit)	expected false discovery rate (FDR) (q) value threshold for Peptide-to-MS2 algorithm, below which the pattern match score is considered significant

There are four files located in *pepgrep* directory, which contain pre-set values for the auto-selection of options required for certain types of MS2 instruments. These files include I.param for ion trap, Q.param for Q-TOF, T.param for triple quad and O.param for Orbitrap instruments. The parameter file can be supplied with argument line using '@' prefix. However, the command line options will override options supplied with the file.

-showgr	uses graphical user interface (GUI) to display matched spectrum
-savegr	saves spectrum plot in a file
-savedta	stores MS2 data of matched spectrum in DTA file

pepgrep on multicore computing environment

The implementation of message passing interface (MPI) to python, *mpi4py* or *pypar* library can be used with *pepgrep*, which is included in *pepgrep.mpi.py*. MPI-*pepgrep* is normally launched from *pepgrep.sh* shell script which invokes the corresponding Open MPI or MPICH2 libraries in a multiprocessor environment. To use the MPI-*pepgrep*, it is necessary to build *mssplitter* binary tool from C++ source files supplied in the corresponding subdirectory. *mssplitter* is invoked before starting *pepgrep.mpi.py* to split the MS2 dataset in parts corresponding to the number of MPI processes. The splitting of the original MS2 spectra file is done according to a spectrum boundary rather than the size of the resulting file. All spectra are indexed in split files to keep the original order of MS2 scans. The command line for MPI-*pepgrep* is similar to single thread *pepgrep*, except that the invocation command line includes shell script *pepgrep.sh*, which is also included in *pepgrep* package.

File formats for MS2 dataset

Four types of most common MS2 formats are supported, including MGF, DTA, PKL and MZXML. File formats are recognized by the file extension.

Peptide-to-MS2 scoring

pepgrep uses Peptide-to-MS2 modelling algorithm for scoring the match between reference and observed patterns in MS2 spectral data for y- and b-ion fragmentation profiles. The module builds simulated MS2 *m/z* ion scans in a form of *numpy arrays* for the reference and decoy peptide sequences at the same time. The decoy sequences are generated using random permutations of the original peptide sequence. The optimal number of decoy sequences used in PMS was found to be equal to 1/3 of number of amino acid residues in the reference peptide. This number was deducted empirically.

The input MS2 spectrum data are parsed into the *numpy arrays* through the iteration of the dataset. The reference and the input array data are compared using *numpy* functions for fast data manipulations in arrays. The matched spectra are scored and the spectrum with the highest rank is subjected to complementary test using MS2 decoy data built with decoy peptide sequences (Figure 1). The score for a tested peptide is calculated using the following equation:

$$S = I_w \times \frac{Y + B}{|Y - B| + 1} \times b \text{ where } I_w = \frac{\sum I_t \times \sum I}{I_m \times L_m \times z} \text{ and } b = 1.0 + \frac{Y_{\text{contig}} + B_{\text{contig}} + YB_p}{L} \quad (1)$$

The scoring formula takes into account the relation to three major properties in potentially matched MS2 scan: (i) cumulative intensity weight (I_w) of the matched ions (adjusted to theoretical intensities), (ii) the total number of y- and b- ions (biased towards the presence of the both types of ions) and (iii) number of scored contiguous ions (b). We refer to I and I_t as the intensity of a peak and theoretical intensity, respectively, which match the corresponding ion from a simulated spectrum. I_m is maximum intensity in the matched spectrum, L_m is the number of matched ions and Z is a charge state. Y and B are the sums of matched y- and b- ions, respectively. Y_{contig} and B_{contig} refer to the number of contiguous y- and b- ions found, respectively, and YB_p is equal to the number of pairs of y- and b- ions supposedly derived from the same fragment product. L is the total number of ions in the model. The range of score values produced is similar to the expectation value produced by TANDEM and Mascot peptide score.

False discovery rate

The statistical significance in Peptide-to-MS2 scoring for both targeted and decoy peptide models was evaluated by implementing the negative binomial distribution model. FDR was calculated for each produced score by invoking complementary cumulative distribution function from *scipy* library with subsequent FDR adjustment to the generated P value. The FDR threshold allows the user to manipulate the significance criteria for selection of the matched MS2 models for output. If FDR for any of MS2 models generated from decoy sequences is less than or equal to a user-defined threshold, then the spectrum is rejected regardless of the search result for the targeted sequence.

Spectrum filtering

pepgrep filters input MS2 data based on certain criteria pertaining to the quality of spectrum. These include such conditions as minimal and maximal mass of the precursor ion, charge state, minimal number of fragment peaks and the ratio between high and low masses in the spectrum.

Spectrum thresholding

The spectrum thresholding includes a single routine for cutting off the peaks that are found below the selected intensity threshold. The intensity threshold can be supplied either with options or otherwise calculated programmatically based on the mass/intensity distribution profile of the spectrum. It is optimized to be used primarily with Q-TOF and Orbitrap types of scans that usually display numerous low intensity 'noise' peaks.

Post-translational modifications

The mass offsets corresponding to various PTMs can be supplied in the file named 'modifications', which can be

found in the *MS* root directory. Practically, the number of PTMs could be limited only by the available system resources. We have found no significant degradation in performance, when *pepgrep* was tested with up to 60 PTMs included in the search using single PC Linux x86_64 system with 16 cores Intel(R) Xeon(R) CPU, 2.40 GHz and 16 Gb memory.

Fixed and non-fixed peptide masses

When ‘non-fixed peptide mass’ is selected as an option, *pepgrep* scores match between MS2 ion fragments only (sequence tags) ignoring peptide masses. On output, it reports the found matches in MS *m/z* data and the calculated difference in mass between predicted and observed peptides. If this option is not selected, then *pepgrep* will try to match both MS2 *m/z* data and peptide precursor masses for the spectrum to be scored.

Selecting Q-TOF-type MS2 by the presence of C-terminal y-ion ladder

Typical Q-TOF-type of CID MS2 spectrum of a tryptic peptide can be easily distinguished from non-peptide spectra by the presence of the y-ion ladder containing first three C-terminal amino acid ions. Thus, a quick test for the presence of the ladder helps to eliminate non-peptide spectra from being processed. This routine is optional and included only in Q-TOF-specific parameters.

Visual support

Matched MS2 spectra can be visualized with *matplotlib* module and/or saved to file for future visual examination.

Competing interests

There are no competing interests to declare.

Acknowledgements

The author would like to thank Dr. Steven Yates for critical overview of the article. This work was supported by the University of Essex.

Supplementary material

Supplementary data associated with this article can be found, in the online version, at <http://dx.doi.org/10.1016/j.gpb.2013.02.001>.

References

- [1] Hunt EF, Yates III JR, Shabanowitz J, Winston S, Hauer CR. Protein sequencing by tandem mass spectrometry. *Proc Natl Acad Sci U S A* 1986;84:620–3.
- [2] Link AJ, Eng J, Schieltz DM, Carmack E, Mize GJ, Morris DR, et al. Direct analysis of protein complexes using mass spectrometry. *Nat Biotechnol* 1999;17:676–82.
- [3] Washburn MP, Wolters D, Yates III JR. Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nat Biotechnol* 2001;19:242–7.
- [4] Shen Y, Tolić N, Xie F, Zhao R, Purvine SO, Schepmoes AA, et al. Effectiveness of CID, HCD, and ETD with FT MS/MS for degradomic-peptidomic analysis: comparison of peptide identification methods. *J Proteome Res* 2011;10:3929–43.
- [5] Dancik V, Addona TA, Clauser KR, Vath JE, Pevzner PA. De novo peptide sequencing via tandem mass spectrometry. *J Comput Biol* 1999;6:327–42.
- [6] Tabb DL, Smith LL, Brezi LA, Wysocki VH, Lin D, Yates III JR. Statistical characterization of ion trap tandem mass spectra from doubly charged tryptic peptides. *Anal Chem* 2003;75:1155–63.
- [7] Perkins DN, Pappin DJ, Creasy DM, Cottrell JS. Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis* 1999;20:3551–67.
- [8] Eng JK, McCormack AL, Yates III JR. An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. *J Am Soc Mass Spectrom* 1994;5:976–89.
- [9] Craig R, Beavis RC. TANDEM: matching proteins with tandem mass spectra. *Bioinformatics* 2004;20:1466–7.
- [10] Sherman J, McKay MJ, Ashman K, Molloy MP. How specific is my SRM?: The issue of precursor and product ion redundancy. *Proteomics* 2009;9:1120–3.
- [11] Addona TA, Abbatiello SE, Schilling B, Skates SJ, Mani DR, Bunk DM, et al. Multi-site assessment of the precision and reproducibility of multiple reaction monitoring-based measurements of proteins in plasma. *Nat Biotechnol* 2009;27:633–41.
- [12] Filippova GN, Fagerlie S, Klenova EM, Myers C, Dehner Y, Goodwin G, et al. An exceptionally conserved transcriptional repressor, CTCF, employs different combinations of zinc fingers to bind diverged promoter sequences of avian and mammalian c-myc oncogenes. *Mol Cell Biol* 1996;16:2802–13.
- [13] Nesvizhskii AI, Aebersold R. Interpretation of shotgun proteomic data. *Mol Cell Proteomics* 2005;4:1419–40.
- [14] Tanner S, Shu H, Frank A, Wang LC, Zandi E, Mumby M, et al. InsPecT: fast and accurate identification of post-translationally modified peptides from tandem mass spectra. *Anal Chem* 2005;77:4626–39.